journal of applied psychology

KENNETH E. CLARK, Editor
University of Rochester

Consulting Editors

GEORGE E. BRIGGS, *Ohio State University*
MARVIN D. DUNNETTE, *University of Minnesota*
NORMAN FREDERIKSEN, *Educational Testing Service*
LEONARD D. GOODSTEIN, *University of Cincinnati*
EDWIN R. HENRY, *University of Rochester*
JOHN HOLLAND, *American College Testing Program*
CLIFFORD E. JURGENSEN, *Minneapolis Gas Company*
DONALD G. MARQUIS, *Massachusetts Institute of Technology*
LAURENCE S. MCGAUGHRAN, *University of Houston*
QUINN MCNEMAR, *University of Texas*
WARREN T. NORMAN, *University of Michigan*
HAROLD F. ROTHE, *Beloit Corporation*
THOMAS A. RYAN, *Cornell University*
ALEXANDER G. WESMAN, *Psychological Corporation*
CLARK L. WILSON, *Instituto Centroamericano de Administracion de Empresas*

Volume 52, 1968

Bureau of Educ. & Psych. Research
(B. C. E. R. I.)
Date 6-3-81
Acc. J 697

HELEN ORR
Managing Editor

MARY C. MCGARRAGHY
Technical Editor

HENRY S. SALMON
Advertising and Business Manager

ROBERT D. GORMAN
Subscription Manager

Published bimonthly by the American Psychological Association, Inc.
Prince and Lemon Sts., Lancaster, Pa. 17604 and 1200 Seventeenth St., N.W.
Washington, D. C. 20036

$10-50

Second-class postage paid at Lancaster, Pa.

© 1968 by the American Psychological Association, Inc.

Contents of Volume 52

ALDERFER, CLAYTON P. **Comparison of Questionnaire Responses with and without Preceding Interviews** 335
ALLEN, THOMAS J. See GERSTBERGER, PETER G.
ALSIKAFI, M., JOKINEN, WALFRID J., SPRAY, S. LEE, AND TRACY, GEORGE S. **Managerial Attitudes toward Labor Unions in a Southern City** 447
ANASTASI, ANNE. See SCHAEFER, CHARLES E.
ATTWOOD, DENNIS A. See WEINER, EARL L.
BABARIK, PAUL. **Automobile Accidents and Driver Reaction Pattern** 49
BACHMAN, JERALD G. **Faculty Satisfaction and the Dean's Influence: An Organizational Study of Twelve Liberal Arts Colleges** 55
BAEHR, MELANY E., & WILLIAMS, GLENN B. **Prediction of Sales Success from Factorially Determined Dimensions of Personal Background Data** 98
BARRETT, GERALD V., & THORNTON, CARL L. **Relationship between Perceptual Style and Driver Reaction to an Emergency Situation** 169
BARRETT, GERALD, V., & THORNTON, CARL L. **Relationship between Perceptual Style and Simulator Sickness** 304
BARRETT, RICHARD S. See KATZELL, RAYMOND A.
BASS, BERNARD M. **Ability, Values, and Concepts of Equitable Salary Increases in Exercise Compensation** 299
BASS, BERNARD M. **How to Succeed in Business According to Business Students and Managers** 254
BASS, BERNARD M. **Interface between Personnel and Organizational Psychology** 81
BEMIS, STEPHEN E. **Occupational Validity of the General Aptitude Test Battery** 240
BESSEMER, DAVID W. See NORMAN, RALPH D.
BLOOD, MILTON R. See NEALEY, STANLEY M.
BOGGS, DAVID H., & SIMON, J. RICHARD. **Differential Effect of Noise on Tasks of Varying Complexity** 148
BORGEN, FRED H. See CAMPBELL, DAVID P. Monograph listing.
BRAUNSTEIN, DANIEL N., & HAINES, GEORGE H. **Preference Scaling of Careers and Organizations** 380
BRAY, DOUGLAS W., & CAMPBELL, RICHARD J. **Selection of Salesmen by Means of an Assessment Center** 36
BRENNER, MARSHALL H. **Use of High School Data to Predict Work Performance** 29
BRIGGS, GEORGE E. See JOHNSTON, WILLIAM A.
BRIGGS, GEORGE E. See NAYLOR, JAMES C.
BROWN, C. HELEN. See POULTON, E. C.
BROWN, EVA METZGER. **Influence of Training, Method, and Relationship on the Halo Effect** 195
BRUVOLD, WILLIAM H. **Scales for Rating the Taste of Water** 245
BRYAN, JUDITH F. See LOCKE, EDWIN A.
BUEL, WILLIAM D. **Evaluating Mortgage Loan Risk** 399
BURG, ALBERT. **Lateral Visual Field as Related to Age and Sex** 10
CAHOON, D. D. **Relative Effectiveness of Programmed Text and Teaching Machine as a Function of Measured Interests** 454
CAMPBELL, DAVID P. See Monograph listing.
CAMPBELL, JOHN P. **Individual versus Group Problem Solving in an Industrial Sample** 205
CAMPBELL, JOHN P. See WERNIMONT, PAUL F.
CAMPBELL, RICHARD J. See BRAY, DOUGLAS W.
CANNON, WILLIAM M. See PERRY, DALLIS K.
CASTORE, CARL H. See STREUFERT, SIEGFRIED.
CHILD, IRWIN L. See KASTL, ALBERT J.
DARLEY, JOHN G. **1917: A Journal is Born** 1
DAVID, KENNETH H. **Age, Cigarette Smoking, and Tests of Physical Fitness** 296
DAVIDSON, P. O. **Validity of the Guilty-Knowledge Technique: The Effects of Motivation** 62
DAWIS, RENE V. See GRAEN, GEORGE B.
DAWIS, RENE V. See THORNDIKE, ROBERT M.
EASTES, SUZANNE H. See CAMPBELL, DAVID P. Monograph listing.
EMANUEL, JOSEPH T. See GOLDSTEIN, IRWIN L.
ERDMANN, R. L., & NEAL, A. S. **Word Legibility as a Function of Letter Legibility, with Word Size, Word Familiarity, and Resolution as Parameters** 403
FISCHL, M. A. See SIEGEL, ARTHUR I.
GERTSBERGER, PETER G., & ALLEN, THOMAS J. **Criteria Used by Research and Development Engineers in the Selection of an Information Source** 272
GHISELLI, EDWIN E. **Interaction of Traits and Motivational Factors in the Determination of the Success of Managers** 480
GIBBS, WILLIAM L. **Driver Gap Acceptance at Intersections** 200
GOLDSTEIN, IRWIN L., EMMANUEL, JOSEPH T., & HOWELL, WILLIAM C. **Effect of Percentage and Specificity of Feedback on Choice Behavior in a Probabilistic Information-Processing Task** 163
GRAEN, GEORGE B. **Testing Traditional and Two-Factor Hypotheses Concerning Job Satisfaction** 366
GRAEN, GEORGE B., DAWIS, RENE V., & WEISS, DAVID J. **Need Type and Job Satisfaction among Industrial Scientists** 286

GRAEN, GEORGE B., & HULIN, CHARLES L. **Addendum to "An Empirical Investigation of Two Implications of the Two-Factor Theory of Job Satisfaction"** 341
GRUENFELD, LEOPOLD W. See WEISSENBERG, PETER.
HAINES, GEORGE H. See BRAUNSTEIN, DANIEL N.
HAMMERTON, M., & TICKNER, A. H. **Comparison of Operator Performance on a Tracking Task Using Discretely and Continuously Variable Controls** 319
HARRIS, DOUGLAS H. **Effect of Defect Rate on Inspection Accuracy** 377
HINTON, BERNARD L. **Environmental Frustration and Creative Problem Solving** 211
HOGAN, JOHN M. See KATZELL, RAYMOND A.
HOLLAND, JOHN L. See Monograph listing.
HOWELL, WILLIAM C. See GOLDSTEIN, IRWIN L.
HULIN, CHARLES L. **Effects of Changes in Job-Satisfaction Levels on Employee Turnover** 122
HULIN, CHARLES L. See GRAEN, GEORGE B.
HULIN, CHARLES L. See MIKES, PATRICIA SMITH.
HURST, PAUL M., PERCHONOK, KENNETH, & SEGUIN, EDMOND L. **Vehicle Kinematics and Gap Acceptance** 321
HURST, PAUL M. See PERCHONOK, KENNETH.
JOHANNSON, CHARLES B. SEE CAMPBELL, DAVID P. Monograph listing.
JOHNSON, PAUL V., & MARCRUM, ROBERT H. **Perceived Deficiencies in Individual Need Fulfillment of Career Army Officers** 457
JOHNSTON, WILLIAM A., & BRIGGS, GEORGE E. **Team Performance as a Function of Team Arrangement and Work Load** 89
JOKINEN, WALFRID J. See ALSIKAFI, M.
KANUNGO, RABINDRA N. **Brand Awareness: Effects of Fittingness, Meaningfulness, and Product Utility** 290
KASTL, ALBERT J., & CHILD, IRVIN L. **Emotional Meaning of Four Typographical Variables** 440
KATZELL, MILDRED ENGBERG. **Expectations and Dropouts in Schools of Nursing** 154
KATZELL, RAYMOND A., BARRETT, RICHARD S., VANN, DONALD H., & HOGAN, JOHN M. **Organizational Correlates of Executive Roles** 22
KENDALL, LORNE M. See LOCKE, EDWIN A.
KIPNIS, DAVID. **Social Immaturity, Intellectual Ability, and Adjustive Behavior in College** 71
KORMAN, ABRAHAM K. **Task Success, Task Popularity, and Self-Esteem as Influences on Task Liking** 484
KUGELMASS, SOL, & LIEBLICH, ISRAEL. **Relation between Ethnic Origin and GSR Reactivity in Psychophysiological Detection** 158
KUSYSZYN, IGOR. **Comparison of Judgmental Methods with Endorsements in the Assessment of Personality Traits** 227
LACHMAN, SHELDON J. **Learning of Telephone Numbers with and without Letter-Exchange Prefixes** .. 95
LANTERMAN, RICHARD S. See SIEGEL, ARTHUR I.
LAWLER, EDWARD E., III. **A Correlational-Causal Analysis of the Relationship between Expectancy Attitudes and Job Performance** 462
LAWLER, EDWARD E., III, PORTER, LYMAN W., & TANNENBAUM, ALLEN. **Managers' Attitudes toward Interaction Episodes** 432
LEVINE, EDWARD L., & WEITZ, JOSEPH. **Job Satisfaction among Graduate Students: Intrinsic versus Extrinsic Variables** 263
LIEBLICH, ISRAEL. See KUGELMASS, SOL.
LIPPITT, GORDON L. See PETERSEN, PETER B.
LOCKE, EDWIN A., BRYAN, JUDITH F., & KENDALL, LORNE M. **Goals and Intentions as Mediators of Effects of Monetary Incentives on Behavior** 104
LUDWIG, D. See MELTZER, H.
MARCRUM, ROBERT H. See JOHNSON, PAUL, V.
MCCONNELL, J. DOUGLAS. **Effect of Pricing on Perception of Product Quality** 331
MELTZER, H., & LUDWIG, D. **Memory Dynamics and Work Motivation** 184
MELTZER, H., & LUDWIG, D. **Relationship of Memory Optimism to Work Competency and Personality Variables** 423
MIKES, PATRICIA SMITH, & HULIN, CHARLES L. **Use of Importance as a Weighting Component of Job Satisfaction** 394
MUKHERJEE, BISHWA NATH. **Achievement Values and Scientific Productivity** 145
MUKHERJEE, BISHWA NATH. **Achievement Values, Social Desirability, and Endorsement of Trait Names on the Berdie Check List** 127
NAYLOR, JAMES C., BRIGGS, GEORGE E., & REED, WALTER G. **Task Coherence, Training Time, and Retention Interval Effects on Skill Retention** 386
NEAL, A. S. See ERDMANN, R. L.
NEALEY, STANLEY M., & BLOOD, MILTON R. **Leadership Performance of Nursing Supervisors at Two Organizational Levels** 414
NORMAN, RALPH D., & BESSEMER, DAVID W. **Job Preferences and Preference Shifts as Functions of Job Information, Familiarity, and Prestige Level** 280
NUGENT, FRANK A. **Relationship of Kuder Preference Record Verification Scores to Adjustment: Implications for Vocational Development Theory** 429
ORNE, MARTIN T. See THACKRAY, RICHARD I.
OSBORNE, FRANCIS H. See TAUB, HARVEY A.
PERCHONOK, KENNETH. See HURST, PAUL M.

PERCHONOK, KENNETH, & HURST, PAUL M. **Effect of Lane-Closure Signals upon Driver Decision Making and Traffic Flow** 410
PERRIMENT, A. D. **Effect of Control Placement on Information Transfer Rate Using Bisensory Signals** 357
PERRY, DALLIS K., & CANNON, WILLIAM M. **Vocational Interests of Female Computer Programmers** 31
PETERSEN, PETER B., & LIPPITT, GORDON L. **Comparison of Behavioral Styles between Entering and Graduating Students in Officer Candidate School** 66
PETERSON, ROBERT A. See CAMPBELL, DAVID P. Monograph listing.
PORTER, LYMAN W. See LAWLER, EDWARD E., III.
POULTON, E. C. **Searching for Letters or Closed Shapes in Simulated Electronic Displays** 348
POULTON, E. C., & BROWN, C. HELEN. **Rate of Comprehension of an Existing Teleprinter Output and of Possible Alternatives** 16
REED, WALTER G. See NAYLOR, JAMES C.
SANDERS, ERIC P. **Evolutionary Performance, Managerial Abilities, and Change: An Exploratory Investigation of Organizations** 362
SCHAEFER, CHARLES E., & ANASTASI, ANNE. **A Biographical Inventory for Identifying Creativity in Adolescent Boys** 42
SCHULTZ, DOUGLAS G. See SIEGEL, ARTHUR I.
SCHUSTER, D. H. **Evaluation of a New Hybrid in Programmed Instruction** 309
SEGUIN, EDMOND L. See HURST, PAUL M.
SEMINARA, JOSEPH L. See SHAVELSON, RICHARD J.
SHAVELSON, RICHARD J., & SEMINARA, JOSEPH L. **Effect of Lunar Gravity on Man's Performance of Basic Maintenance Tasks** 177
SIEGEL, ARTHUR I., SCHULTZ, DOUGLAS G., FISCHL, M. A., & LANTERMAN, RICHARD S. **Absolute Scaling of Job Performance** 313
SIMON, J. RICHARD. **Reaction Time as a Function of the Cue Properties of an Auditory Display** 224
SIMON, J. RICHARD. See BOGGS, DAVID H.
SMITH, PATRICIA CAIN. See ZEDECK, SHELDON.
SMITH, ROBERT D. **Heuristic Simulation of Psychological Decision Processes** 325
SPRAY, WALFRID J. See ALSIKAFI, M.
STREUFERT, SIEGFRIED, STREUFERT, SUSAN C., & CASTORE, CARL H. **Leadership in Negotiations and the Complexity of Conceptual Structure** 218
STREUFERT, SUSAN C. See STREUFERT, SIEGFRIED.
STUDENT, KURT R. **Supervisory Influence and Work-Group Performance** 188
TANNENBAUM, ALLEN. See LAWLER, EDWARD E., III.
TAUB, HARVEY A., & OSBORNE, FRANCIS H. **Effects of Signal and Stimulus Rates on Vigilance Performance** 133
THACKRAY, RICHARD I., & ORNE, MARTIN T. **Effects of the Type of Stimulus Employed and the Level of Subject Awareness on the Detection of Deception** 234
THORNDIKE, ROBERT M., WEISS, DAVID J., & DAWIS, RENE V. **Multivariate Relationships between a Measure of Vocational Interests and a Measure of Vocational Needs** 491
THORNTON, CARL L. See BARRETT, GERALD V.
TICKNER, A. H. See HAMMERTON, M.
TRACY, GEORGE S. See ALSIKAFI, M.
VANN, DONALD H. See KATZELL, RAYMOND A.
WEINER, EARL L., & ATTWOOD, DENNIS A. **Training for Vigilance: Combined Cueing and Knowledge of Results** 474
WEISS, DAVID J. See GRAEN, GEORGE B.
WEISS, DAVID J. See THORNDIKE, ROBERT M.
WEISSENBERG, PETER, & GRUENFELD, LEOPOLD W. **Relationship between Job Satisfaction and Job Involvement** 469
WEITZ, JOSEPH. See LEVINE, EDWARD L.
WERNIMONT, PAUL F., & CAMPBELL, JOHN P. **Signs, Samples, and Criteria** 372
WILLIAMS, GLENN B. See BAEHR, MELANY E.
ZEDECK, SHELDON, & SMITH, PATRICIA CAIN. **A Psychophysical Determination of Equitable Payment: A Methodological Study** 343
ZURCHER, LOUIS A. **Particularism and Organizational Position: A Cross-Cultural Analysis** 139

MONOGRAPHS

BORGEN, FRED H. See CAMPBELL, DAVID P.
CAMPBELL, DAVID P., BORGEN, FRED H., EASTES, SUZANNE H., JOHANNSON, CHARLES B., & PETERSON, ROBERT A. **A Set of Basic Interest Scales for the Strong Vocational Interest Blank for Men.** No. 6, Part 2.
EASTES, SUZANNE H. See CAMPBELL, DAVID P.
HOLLAND, JOHN L. **Explorations of a Theory of Vocational Choice: VI. A Longitudinal Study Using a Sample of Typical College Students.** No. 1, Part 2.
JOHANNSON, CHARLES B. See CAMPBELL, DAVID P.
PETERSON, ROBERT A. See CAMPBELL, DAVID P.

journal of applied psychology

Vol. 52, No. 1, Part 1 February 1968

1917: A JOURNAL IS BORN

JOHN G. DARLEY

University of Minnesota

1967 represented the 50th yr. of publication of this Journal. A reading of 10 complete volumes on a quinquennial cycle starting in 1917 suggests that the classic problems of applied psychology were approximately adumbrated by about 1922, except for the area of biomechanics. Many of today's fashionable concepts, such as moderator variables, cross-validation, question format, are easily discernible in early volumes. The disappearance of several early tests as the technology improved is noticeable. Judging by bibliographic citation patterns, applied psychologists share psychologists' generally cavalier attitude toward history. Thus this excursion into the history of 1 journal may commend itself to other amateur historians as an interesting and possibly worthwhile didactic exercise.

The time was March 1917. The United States was engaged in the "war to end war"; it came to be known later as World War I. Volume 1, Number 1, of the *Journal of Applied Psychology,* Florence Chandler of Worcester, publisher, went to a subscription list of now unknown size. Its foreword by G. Stanley Hall, John Wallace Baird, and L. R. Geissler described its goals:

> The past few years have witnessed an unprecedented interest in the extension of the application of psychology to various fields of human activity. . . . Thence the attention of the applied psychologist turned to the more comprehensive and fundamental problem of vocational selection . . . and it must appeal to every human being who is interested in increasing human efficiency and human happiness by the more direct method of decreasing the number of cases where a square peg is condemned to a life of fruitless endeavor to fit itself comfortably into a round hole . . . none of the existing journals devote themselves to the task of gathering together the results of workers in the various fields of applied psychology, or of bringing these results into relation with pure psychology. . . . In view of all these circumstances the undersigned have decided, not without hesitation, to launch a Journal of Applied Psychology,—their hesitation being of financial origin because they have no pecuniary resources in hand or in prospect, excepting their own . . . [pp. 5–7].

The masthead identified the cooperating editors; these were the first of many psychologists who served effectively in this anonymous but crucial gate-keeping role: W. V. Bingham, June E. Downey, Robert H. Gault, Dr. William Healy, H. L. Hollingworth, Joseph Jastrow, Herbert S. Langfeld, James H. Leuba, Henry C. McComas, Eleanor Harris Rowland, Walter D. Scott, C. E. Seashore, Daniel Starch, Edward K. Strong, Jr., L. M. Terman, E. L. Thorndike, J. E. Wallace Wallin, G. M. Whipple, Helen Thompson Woolley. The masthead goes on further to say:

> The field of the Journal will include the following:
>
> (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeal.
>
> (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis.
>
> (c) The influence of general environmental conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc.
>
> (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or musical instruments, sports, and the like.
>
> In reply to several inquiries the editors wish to state that those mental tests which aim to deal with

© 1968 by the American Psychological Association, Inc.

purely educational problems or abnormal minds do not as a rule fall within the scope of the JOURNAL [pp. 1–2].

The contributors were expected to pay certain costs but "as soon as the JOURNAL becomes self-supporting, or as soon as the present excessive prices of paper and printing abate, the above terms will be made more liberal [p. 2]." The record indicates that the excessive prices did not abate.

In his autobiography, Hall (1923) makes only the most casual reference to this new Journal. Writing about the evolution of the testing movement as an outgrowth and extension of one of his major loves—child study—he says:

It is in the great stress now laid upon these tests, too, that psychology in general has enormously increased its utility, but not without grave danger of lapsing from culture to Kultur, since the later methods seek practical results with individuals and have added little to our knowledge of child nature in general. In this latter direction, however, Clark University has made it a point for the last decade or more to have some competent authority to apply tests gratis to all children from the city sent or brought to it, and in recent years Dr. L. R. Geissler gave his entire time to making such tests and instructing others to do so here, and we founded here the Journal of Applied Psychology, an organ for the development of this work which, as I write, is in its sixth volume [p. 399].

This is the only mention of Geissler in the autobiography; there is no mention whatever of Baird, the other coeditor. Baird had been at Clark since 1910; he had been in charge of the laboratory there and in 1918 served as president of the American Psychological Association. He died in 1919, having served as coeditor for only 2 years.

Hall was far more enamored of the *American Journal of Psychology,* which he founded in 1889, and the *Pedagogical Seminary,* which he founded in 1893. In another place in his autobiography, he speaks more harshly of psychometrics:

We have no method of testing testers, whose competence varies vastly from mere mechanical accumulation of data and the busy work of computing correlations up to real insight and sagacity. . . . I would cast no discredit upon this work, much as I deplore its present monopoly of psychological endeavor, but to my mind no one has made it sufficiently clear just what even the most used of them really tests, and still less justified for either psychology or for life the value of the powers they activate. All of them together have done a great work in applying psychology to life and industry but have added scarcely a scintilla to our knowledge of the human soul [p. 450].

Editors may propose, but authors and the availability of manuscripts dispose. There were 35 articles, 37 authors, and 399 pages in Volume 1. The volume also included book reviews and notices, brief notes and summaries, and a section on psychology and military affairs covering the deployment of psychologists, the work of special committees, and other special activities related to World War I, signed by Robert M. Yerkes, Major, Sanitary Corps, and Chairman Psychology Committee, National Research Council.

Using the editors' four fields for the new Journal, these 35 articles are classified in the 1917 column of Table 2. It is evident that contributions were accepted under less than rigorous criteria in the first year. The essays ranged from Hall's own APA twenty-fifth anniversary speech on psychology and the war through two philosophic articles by A. A. Roback, in one of which he attacks Munsterberg's avoidance of moral and ethical issues in applied psychology. The doubtful felicity of this critique lies in the fact that William Stern's memorial tribute on the occasion of Munsterberg's death is found in the same volume. Other essays deal with applications of psychology in a wide range of endeavor—with its utility to business men, its promise of solutions to the problems of labor turnover, its value in increasing the efficiency of workers. In the main, these are characterized by a high moral tone, by hope rather than evidence, and by a very conservative economic outlook.

By today's standards, statistical analyses were primitive and often missing. A few articles, however, are sophisticated in design, analysis, or presentation. Baird's study of the legibility of telephone directories in various type faces; Burtt's analyses of Munsterberg's tests applied to various samples; Miner's correlational study of rating methods; Pintner and Toops' test analyses of unemployed men; Scott's presentation of test results of applicants versus successfully employed as a validating method; Terman's use of psychological tests in selecting policemen and firemen; Thorndike's study of decrements in work output and work satisfyingness; Yerkes' com-

parison of Binet and point scale measures of intelligence—each of these is a delight to read. Wallin, in a long case history of a stutterer, takes a direct slap at Freudian theory:

> But to maintain that all the words in a custom-made list which are associated with delayed responses reveal an inner conflict or emotional complex, may suffice for the unsophisticated devotee of a cult but surely cannot satisfy the critical experimental scientist who demands facts capable of objective verification [pp. 358–359].

Garrison, Burke, and Hollingworth present a case report on a superior child first tested at 8 years and 4 months of age, with a Stanford-Binet of 187; follow-up reports on this child appear with pleasant regularity in Volumes 6 and 11 of the Journal.

And so the new journal was launched. It took its place in brilliant company: *Psychological Review* was, in 1917, 23 years old and its monograph supplements first appeared in 1896; the *Journal of Experimental Psychology* had been established in 1916; the *Psychological Bulletin* was 13 years old; the *Psychological Index* had been in existence since 1894; the *Journal of Abnormal Psychology* first appeared in 1906; the *Journal of Educational Psychology* was first published in 1910; *Psychological Monographs* first appeared in 1910; G. Stanley Hall's first love, the *American Journal of Psychology,* started publication in 1889, the year in which Hall opened, as president, the doors of Clark University.

To see how the journal evolved, I have read Volumes 1, 6, 11, 16, 21, 26, 31, 36, 41, and 46 in their entirety, thus sampling almost a half-century of the Journal's life. Some of the material I have tabled. (See Tables 1, 2, and 3.) At other points I have commented on trends or items of interest—to me at least.[1]

The chronology of editors and ownership deserves brief comment. In 1921, James P. Porter, Clark University, and William F. Book, Indiana University, appeared as editors. Hall had apparently sold the Journal in 1920 or 1921; his ardor for it, certainly, had cooled. Porter left Clark to go to Ohio University; in 1922 his new affiliation first appeared. By 1928, Porter was the sole editor; Book apparently withdrew in 1927. The *Psychological Bulletin* for 1943 (Volume 40, page 652), reporting the annual business meeting of the APA, noted that the Journal was purchased from James Porter that year. Volume 27, 1943, was edited by Donald G. Paterson and published by the APA with the cooperation of the American Association for Applied Psychology. In 1955, Volume 39, I started a 6-year term as editor. In 1961, Volume 45, Kenneth E. Clark became the editor and continues today in that position. Porter's tenure lasted 22 years; Paterson held the editorship for 12 years. The years of the Journal's existence have been the years of psychology's most rapid growth. New journals have come into being in the area of counseling psychology, in the areas of psychometrics and testing, the area of human engineering, and the area of personnel management, to mention some of the fields for which the *Journal of Applied Psychology* has provided a publication outlet.

TABLE 1

NUMBERS OF ARTICLES, PAGES, AND AUTHORS IN SELECTED VOLUMES OF THE JOURNAL OF APPLIED PSYCHOLOGY

Year	Volume number	No. articles	No. pages	No. authors	No. early publications
1917	1	35	399	37	—
1922	6	33	433	37	—
1927	11	45	544	51	—
1932	16	68	717	81	—
1937	21	59	735	66	—
1942	26	81	887	102	—
1947	31	80	664	96	27
1952	36	97	433	132	21
1957	41	89	418	131	4
1962	46	85	432	141	7

Note.—Book reviews, notes and news, materials received appeared through Volume 39. *Contemporary Psychology,* established in 1956, removed these pages from the *Journal of Applied Psychology.*

Let us go back, now, to our quinquennial sampling of the volumes. Volume 6 covers the year 1922. There were 33 articles, 433 pages, and 37 authors. The number of dataless essays has dropped to two; four articles that must be classified as statistical or methodological now appear: Toops described a method of computing a matrix of intercorrelations by using an adding machine; Hull described a

[1] The structure of this review owes much to David Cohn's (1940) *The Good Old Days,* a fascinating and exciting treatment of our social history. I regret my lack of time and skill to duplicate Cohn's analysis.

TABLE 2

PERCENTAGES OF ARTICLES IN VARIOUS CATEGORIES OR FIELDS

Year	Volume number	Applications to vocational activities	Studies of individual mentalities	Influence of environmental conditions	Psychology of everyday activities	Essays	Case studies	Statistical/methodological
1917	1	20	26	0	6	36	9	3
1922	6	27	49	0	3	6	3	12
1927	11	22	47	0	13	4	7	7
1932	16	26	47	2	13	0	2	10
1937	21	26	34	3	20	7	3	7
1942	26	31	32	2	12	5	4	14
1947	31	44	21	5	9	7	0	14
1952	36	50	18	10	7	8	0	7
1957	41	27	25	26	16	0	0	6
1962	46	34	25	27	1	0	0	13

Note.—It was immediately evident in Volume 1 that the editors' original four fields were insufficient; thus these other categories of articles had to be introduced in the classification scheme. Additionally, I have no evidence on the reliability or interjudge consistency of the categories, since I was the sole judge of each article's place in the scheme.

method of deriving sigma scores, and a method of computing Pearsonian r's from ranked data; Thorndike discussed methods of deriving equivalent scores on intelligence tests of varying degrees of correlation, means, and dispersals; Chapman discussed the problem of known and guessed scores in true-false examinations.

Ream, in an article describing the development of a test to measure Thorndike's "social intelligence," mentions socially acceptable and unacceptable items, foreshadowing the later spate of literature on item format. Poffenberger concludes an article on setting cutting scores with the following statement:

> From the psychologist's point of view it might not be difficult to decide upon a critical intelligence score for the purposes of this school. . . . But here as in most institutions where tests are being established the keenest interest centers upon the individual case. If one or a few good students be lost, the test may thereby be condemned. "If a man have a hundred sheep and one of them be gone astray, doth he not leave the ninety and nine, and goeth into the mountains, and seeketh that which is gone astray? And if so be that he finds it, . . . he rejoiceth more of that sheep than of the ninety and nine which went not astray." [p. 160]

So much for selection ratios, institutional decisions, and predictive validity. Poffenberger, in another article, uses peer ratings as an independent variable in predicting the criterion of expert ranks of letters of application. Max Freyd introduces the methods later used by Strong in the measurement of vocational interests. Root, in one of the two essays, gives a sophisticated discussion of the

TABLE 3

PERCENTAGES OF AUTHORS IN VARIOUS TYPES OF EMPLOYMENT

Year	Volume number	Consulting or industry	University or college	Military service	State/city service	Federal service	Private practice	Not known
1917	1	14	66	0	8	0	6	6
1922	6	3	73	0	21	0	0	3
1927	11	6	77	0	11	0	0	6
1932	16	2	72	0	15	3	3	6
1937	21	9	76	0	15	0	0	0
1942	26	19	60	4	13	0	0	4
1947	31	30	62	4	2	1	1	0
1952	36	10	79	5	2	2	2	0
1957	41	16	61	20	1	1	0	1
1962	46	18	59	20	1	2	0	0

clinical versus the statistical problems underlying the battle over the constancy of the IQ. Goldsmith reports the complete standardization of a weighted application blank for use in predicting success of life insurance salesmen. Naccarati and Lewy-Guinzberg, writing on the relation between hormones and intelligence, display various correlations between height-weight ratios and intelligence test scores, explaining away the low correlations by sampling errors that led to the inclusion of too many normosplanchnics in the student samples, at the expense of microsplanchnic and macrosplanchnics. J. McKeen Cattell announces the availability of shares of stock in the newly organized Psychological Corporation at $10 per share. Knight shows an amazing mean difference between teacher volunteers who signed their names and those who did not in scores on a scholastic ability test in a program introducing them to the idea of mental testing. Viteles compares the intercorrelations and differential predictive power of three different intelligence tests, using college grades as criterion measures. His findings indicate that tests of presumably the same functional domain do not show high intercorrelations or equal predictive power. One gets the feeling, from Volume 6, that quite a few of the classic problems of applied psychology had been opened up fairly early.

In Volume 11, 1927, three methodological articles discuss partial correlations, weighting and scaling methods for mental tests, and theorems concerning the arithmetic mean and standard deviation. Kornhauser, standardizing a questionnaire to differentiate between good and poor students at the University of Chicago, clearly describes the procedure of cross-validation, although he does not so name it:

> The unreliability of the questionnaire results became further apparent through the following comparison. In order to check the results obtained from our sample of 45 questionnaires, another sample of 30 . . . was taken. Of the 30 questions which gave the best differentiation in the first sample, only eight gave even fairly good results in the new sample. The correlation of total scores on the 30 items, with average marks, was 0.20 in this group of 30 people, as contrasted with the 0.73 in the original sample [p. 93].

Reporting on two case histories of a pair of sisters placed in an orphanage setting, Florence Teagarden deals with the nature-nurture controversy as follows:

> By way of conclusion it may be said that the improvement in moral, physical, and domestic habits and skills seems to justify the move that was inaugurated for these two children. Their quality of intelligence, however, has remained unchanged during six years. They were feeble-minded in the old environment and they are feeble-minded in the new environment. Finally it may be said that their old environment is probably only an index of the intelligence of the stock from which the children come [p. 296].

So much for nurture!

In two articles, South deals with psychological aspects of committee work: "The committee is a perquisite of democracy . . . today a goodly part of the world's deliberations takes place in committees [p. 348]." His research designs deal with committee size, type of problem, sex of committee members, role of intelligence, and time available for committee activities. Considering that these questions are still not clearly understood by our colleagues in social psychology, it is interesting to see such an early applied approach to these problems. The moral tone still sounds occasionally in Volume 11. In a job analysis of athletes, Charles Homer Bean writes:

> Educators seem to be no longer in doubt about the value of games and other athletic sports, not only for health and for the consumption of surplus energy that would otherwise bother the teacher, but also for the education of both body and mind. The boy or girl who enjoys few, or no opportunities for vigorous athletic training seems to grow to adulthood unable to use, without confusing amounts of static, the various psychoneural hook-ups that are as all-important to clear thinking as to the efficient converting of thought into action [p. 369].

The notes and news columns of Volume 11 contain the following two items:

> "Motion picture theaters are too dark," Guy A. Henry, of New York, General-Director of the Eye Sight Conservation Council of America, declared in a recent statement. "People are unnecessarily subjected to eye strain," he asserted, "in poorly lighted auditoriums." Investigations reveal that managers of motion picture theaters have no method of determining the effectiveness of the lighting, said Mr. Henry, who urged the framing of a special code of illumination for motion picture auditoriums, following a scientific study of the problem. He believes that this is too important a matter to be left to the judgment of individuals [pp. 165–166].

> Defective eyesight contributes to juvenile delinquency, according to Guy A. Henry, General-Direc-

tor of the Eye Sight Conservation Council of America. Child Health Day, general observance of which has been asked in proclamations by mayors of cities and other officials throughout the country, should be made an event of moral and physical significance in every household where there are children.

Investigation by the Eye Sight Council has shown that "bad eyes make bad boys," and that the experience of Juvenile Courts prove that defective vision makes children truants. Each year more than 200,000 children come before these courts, and eye conservation in the home and in the schools should be employed as a factor in checking criminal tendencies.

"Juvenile Courts are finding that bad eyesight leads to inattention in school, to unfair competition, and to disrespect for authority. Bad eyes lead to truancy and the truant child is a criminal in embryo." [p. 319]

The editorial standards for notes and news were, apparently, not stringent.

Volume 16 showed noticeable increases in total pages, number of articles, and number of authors. However the proportions of articles in various fields of applied psychology did not change markedly.

Two items from the Notes and News sections of Volume 16 are interesting:

According to a recent report by the New York Commission on Ventilation there are still twenty states in the Union which retain statutes concerning ventilating devices which are based on disproved or antiquated theories. Dr. C.-E. A. Winslow, chairman of the Commission, estimates that an unnecessary expenditure of nearly $2,500,000 annually is made in an attempt to conform to obsolete regulations requiring the installation and maintenance in school buildings of a system of mechanical ventilation which science has proved to be not only unnecessary but even a menace to health [p. 97].

Syracuse University announces the establishment of an independent department of psychology, with Dr. Ernst Thelin as head of the department. Until the present time, psychology and philosophy have been operating as a single department at Syracuse, despite the fact that in nearly all other universities they have established separate organizations. At Syracuse such a change was deferred, owing to the satisfactory operation of both these branches of teaching under the head of the Department of Philosophy. But in recent years psychology has grown at Syracuse from a simple department in Philosophy to a group including 16 sections of Introductory Psychology and 18 advanced courses, with a total enrollment of 1,000 students [p. 98].

The only case study reported in Volume 16 is S. H. Britt's "Examination of Miss Gene Dennis, 'Psychic'." Miss Dennis was apparently laying them in the aisles in St. Louis in her theatrical performances as *"a real seer* [p. 691]." Dr. Britt, having examined her at her own request, states:

The experiment seems to justify the following conclusions with regard to Miss Dennis:

1. In terms of IQ, she was slightly below average.
2. In the solution of a learning problem, she was considerably less efficient than the *average* college student.
3. As to "mind-reading" ability, she failed to demonstrate any.
4. In no part of the experiment was there evidence of "psychic" powers [p. 694].

There is no evidence of the effect of these conclusions on Miss Dennis' subsequent career.

Volume 16 carries an excellent study of faking of vocational interest test scores by H. C. Steinmetz; among his suggestions: "Further experimentation of this sort with personality tests is suggested for measurement of social insight and the extent to which purpose may distort test results [p. 130]."

An interesting observation in Volume 16 on something akin to socialized medicine is found in the following introduction to an article by Keller and Weber:

At present all persons seeking to prepare for the dental profession in Germany must take and successfully pass a battery of searching tests. Before proceeding to an account of the tests it will be well to say something regarding the training and status of the dentist in Germany. The dentist is not to be confused with his academic competitor, the *Zahnarzt*. The *Zahnarzt* is fully educated in the academic sense and holds a full medical degree. He is a medical doctor, perhaps also a surgeon, who has specialized in diseases and troubles of the teeth. The dentist on the other hand is a practical though perhaps highly skilled technician. We are speaking here of the *Staatlichgeprueften Dentisten* who form the National Association of Dentists (*Reichsverband Deutscher Dentisten*) with head offices in Berlin. Among others these dentists are permitted to take as patients the large number of people who are insured in the *Krankenkasse,* one of the great social institutions of Germany. Certain classes of people, particularly laborers, are required to pay small fees to the *Krankenkasse* as a premium for medical services to which they are entitled when need arises. This system has been found successful in combating the tendency which is world-wide to neglect ills—naturally one takes advantage of medical services for which one has already paid [p. 465].

Their evidence on validation is given briefly: "It has never occurred that a candidate who was found 'suitable' in the test was afterwards found unsuitable in college work, ex-

cept in one case in which, however, a psychosis was involved [p. 474]."

Volume 16 seems to have a report on the first use of subliminal stimulation affecting consumer behavior. Entitled "How the Consumer Estimates Quality by Subconscious Sensory Impressions," the article describes a field test of four identical pairs of women's stockings treated with traces of four different scents. The author blandly quotes from a local newspaper the reason for discontinuing the field work:

> Police learned the efforts were made at the suggestion of Prof. Donald Laird, professor of psychology at Colgate. They reported that the experimenters carried a box of three pairs of stockings which were perfumed and another with stockings that were without the artificial odor and their errand was to learn whether the housewives would consider the perfumed article more attractive than those without perfume [p. 245].

Laird's own conclusions:

> Two things are plainly indicated by this experiment: 1. Subconsciously perceived sensory impressions received through the olfactory apparatus are potent in determining the housewife's judgment of quality in silk hosiery. 2. Scents of one type are more influential in determining this judgment of quality than are scents of another type [p. 246].

Applications of psychophysical methods are not missing in Volume 16. A modest study by G. R. Wendt, then of the Yale Institute of Human Relations, presents a modest summary:

> The Method of Constant Stimuli was applied to two problems in the measurement of psychological factors in industrial products. Using this method it was shown that a rubber tile floor composed of alternate light and dark squares will appear brighter when its pattern is made up of larger units. The amount of this apparent brightness difference was measured. In a second application of the same method it was found that the character of cover markings of golf balls has no effect on their apparent size.
>
> Such applications of experimental psychological methods seem to the writer to offer a wide and as yet untouched field for work in applied psychology. The applications of psychology to industry have included personnel studies, studies relating to the adjustment of the worker to his job and studies in the psychology of advertising. Except for an occasional study of consumers' preferences little measurement of psychological factors in industrial products has been attempted. The field is one which recommends itself not only because of its range of possibilities but also because of the relative ease with which its problems can be solved [p. 276].

Volume 21 finally contains an article *clearly* falling in the psychology of everyday life; Henry Adams provides a statistical analysis of golf. Several laws derive from this study:

> Other things being equal, the average number of strokes needed to play a hole is a straight line function of the length of the hole. . . . The average effective length of the drive of the composite tournament winner is 228.96 yards. . . . The effective length of the full second shot as played by the tournament winners is 228.96 yards. . . . The humanly perfect shot is one having a deviation of approximately 5° on both sides of the intended line of flight of the ball [p. 384].

These "laws" are cold comfort to weekend golfers.

Three articles in this volume deal with automobile driving and traffic problems, the first such articles in this sampling of the Journal. Andrew presents the first factor analysis seen in this quinquennial review, in her report on the standardization of the Minnesota Vocational Test for Clerical Workers.

In a study of the job satisfaction of psychologists, Hoppock sampled the APA membership in vocational or industrial psychology; out of 203 job satisfaction questionnaires mailed, 66 psychologists responded. "During the four years from 1932 to 1936 the average salary had decreased $157 a year to a 1936 average of $3,261 [p. 301]." The good old days!

The formation of the new American Association for Applied Psychology is mentioned in the Notes and News columns and in a more extended article by Fryer, following upon the APA meeting in Minneapolis in 1937.

Another excellent example of cross-validation, though not so called, appears in the article by Helen Messenger entitled: "An Illustration of New Techniques in Test Construction [p. 311]." A little-known and, I must admit, long-forgotten article by Darley (page 485) bears a fair resemblance to what later came to be called studies of moderator variables; a similar study of moderator variables, also affecting the prediction of academic success, is presented in a study by Wolf (page 304).

Volume 26, 1942, marked the last year of Porter's editorship. It contained 81 articles and 887 pages. The first appearance of the

analysis of variance, in this quinquennial review, is found in two articles; one, by Brown, Fisk, and Torkelson, reports on recovery from glare blindness under conditions of changes in dietary intake of Vitamin A; the other, by Jackson, Jerome, and Schoenfeld, reports a study of the effectiveness of deodorant creams!

World War II makes its presence felt in Volume 26 in the Notes and News columns; in one article on Naval recruit testing by Lewinski (page 316); in an essay by English on the applications of psychology to problems of military training (page 3); in Link's report of attitude surveys of workers in a war economy; in Ruch and Young's analysis of the effectiveness of Axis propaganda; and in an early and favorable book review of Tolman's *Drives Toward War*. Psychology's response to World War II reached journal publication somewhat more rapidly than was the case in World War I.

Another excellent example of cross-validation is found in the article by Ryan and Johnson (page 543) on the use of interest measures in the selection of salesmen and servicemen.

Volume 31, 1947, is the first volume sampled here under Paterson's editorship. As Table 1 will show, 27 articles (or 34% of all articles in the volume) represented early publication paid for by the authors. One gets the feeling the option was somewhat overused.

The first extended discussion of the field of human engineering or biomechanics appears in Volume 31; Dunlap's presidential address to the Division of Consulting Psychology at the Denver APA meetings (1947) describes the wartime emergence of the field and predicts additional areas of application in forthcoming years. As will be seen in Table 2, the little used category of "influence of environmental conditions," in Volume 31 and subsequent volumes under review, becomes heavily loaded with human engineering articles.

Clark, in the midst of many favorable reports of military applications of psychology in these volume years, takes from Navy reports examples of projects that were of limited value or were total failures in contributing to military efficiency (page 94). This article will be required reading at the start of World War III.

Volume 36, 1952, gives a dramatic illustration of editorial power; as Table 1 indicates, it contained 97 articles, the largest number in the volumes reviewed here but packed within only 433 pages that also included book reviews and other materials. Since the proportion of these articles in the various areas of applied psychology did not differ markedly from other years, readers were merely exposed to more of the same diet, with the notable exception of a decrease in articles about advertising research. One of these articles (Twedt, page 207), however, applies factor-analytic methods of mechanical and content variables in ads as predictors of a readership criterion; it is the most effective theoretical study of advertising to be found in the volumes reviewed.

A pleasant air of controversy appears in Volume 36; Eells and Darley exchanged rather brisk comments on Darley's unfavorable review of Eell's coauthored volume on intelligence and cultural differences; Farr, Jenkins, Paterson, and England exchange bland rebuttals with Klare and with Flesch regarding their earlier article on the simplification of Flesch's reading ease formulas. Bernard criticizes Rock's "Sales Situation Test"; Rock answers Bernard's criticism, and Paterson replies to Bernard's passing comment that he shouldn't have published Rock's original article in the first place! With regard to industrial applications of psychology, this volume contains examples of more sophisticated studies of industrial organizations, as distinct from prediction studies of performance of personnel. An article by C. G. Brown and B. J. Neitzel, entitled "Communication, Supervision, and Morale," illustrates this shift in emphasis (page 86).

In Volume 41, there is a noticeable increase in the number of articles classified as human engineering (in the area, "influence of the environment"). I was serving my third year as editor; APA policies had set an allowance of approximately 400 pages per volume year for this Journal. I seemed to have continued Paterson's emphasis on *short* articles, or at least contributors had continued to submit shorter articles after Paterson's terms as editor. A somewhat increased level of sophistication in choice of problems and applications of methods seems to appear in this volume, but I hasten to disclaim any editorial credit for this;

sooner or later the field simply *had* to become more sophisticated!

By Volume 46, 1962, Kenneth E. Clark had assumed the editorship of the Journal. In Volume 41 and again in Volume 46, an increase in the proportions of authors from other than academic employment may be seen, reflecting in part the changing nature of employment patterns in psychology. (See Table 3.)

My colleagues in history find themselves ill at ease in the years of the present; they can neither identify trends nor evaluate factors producing change. I find that I share their malaise as I move to the present years of the *Journal of Applied Psychology*. It would be well, therefore, to draw the curtain of this review with Volume 46, 1962.

But what of the past and what, in broad outline, can be learned from it? Several general observations come to mind as I return these 10 dusty volumes to their dusty shelves.

Plus ça change, plus c'est la même chose. There is obviously no domain of knowledge or unitary field of psychology that can be subsumed under the word "applied"; rather there are many ways and places to apply psychology. In large part, these were nearly all foreshadowed in the early volumes of this Journal. Today's concepts of cross-validation, moderator variables, and question format, for example, are easily discernible in early articles. As Xhignesse and Osgood (1967) have recently pointed out in studying the time dimension of journal citations,

> . . . the time perspective of psychology is short and getting shorter. If this is true, then it would seem that psychology as a science is in danger of "forgetting where it has been," of repeatedly rediscovering facts and theories that have been well worked in the past [p. 790].

Certainly applied psychologists seem to have a cavalier attitude toward their own history. And, as Santayana pointed out, those who forget their own history seem doomed to relive it; there is much that is repetitious in the enterprise.

At the same time, there was and still is an air of excitement and discovery as one follows the *Journal of Applied Psychology* over time, seeing on occasion a new application, an interesting application, or an exciting methodological solution to a persistent problem in human affairs. That such articles are rare or infrequent is not, I submit, characteristic of this Journal alone; the truly seminal or important articles in any area of psychology are infrequent events.

Another question: where are the tests of yesteryear? Many tests appear in early articles that by now have disappeared into well-deserved oblivion, as the psychotechnology and standards of test construction have become more sophisticated and rigorous. Hall did not foresee this, nor did he foresee the contribution that test developments and statistics would make to our understanding of the organization of behavior. The *Journal of Applied Psychology* has always been an outlet for this area, for strong and weak articles.

In the main, authors in this Journal have been writing about performance differences related to age, sex, race, special group membership, or (rarely) special conditions set by an experimenter. These studies, in Cronbach's (1957) definition of the two sciences of psychology, were concerned with events that nature herself had defined and created; they also, in more modern terms, dealt both with construct validity and predictive validity. In any new science, many such studies are needed to clear away the underbrush of "common sense," to test assumptions of causality or covariation, to develop methods and experimental expertise, and to open up new areas of study in the field and in the laboratory. The pages of the *Journal of Applied Psychology* reflect this evolution, with both its strength and its weaknesses, its redundancy, and its infrequent points of high discovery.

REFERENCES

COHN, DAVID L. *The good old days: A history of American morals and manners as seen through the Sears Roebuck catalogues 1905 to the present.* New York: Simon and Schuster, 1940. xxxiv plus 597 pages.

CRONBACH, L. J. The two disciplines of scientific psychology. *American Psychologist,* 1957, **12**, 671–684.

HALL, G. S. *Life and confessions of a psychologist.* New York: Appleton, 1923. ix plus 623 pages.

XHIGNESSE, L. V., & OSGOOD, C. E. Bibliographic citation characteristics of the psychological journal network in 1950 and in 1960. *American Psychologist,* 1967, **22**, 778–791.

(Early publication at request of Editor)

Journal of Applied Psychology
1968, Vol. 52, No. 1, 10-43

LATERAL VISUAL FIELD AS RELATED TO AGE AND SEX [1]

ALBERT BURG

Institute of Transportation and Traffic Engineering
University of California, Los Angeles

Lateral nasal and temporal visual field measurements were obtained by means of a screening perimeter for nearly 17300 Ss, ages 16-92. The major findings are: (a) temporal and total fields are maximum to about age 35, after which field constricts progressively with advancing age, (b) nasal fields increase to a maximum occurring about age 35 or 40, after which a progressive decline takes place, and (c) females consistently demonstrate slightly larger visual fields than men. Possible interpretations for these and other findings are presented, and additional research is suggested to explain some of the relationships obtained in the study.

From 1961 through 1966, the author directed a large-scale research project investigating the relationships between vision and driving. The major conclusions from this study are presented in a recent report (Burg, 1967b).

Apart from data relating vision to driving, the study produced a very great amount of information concerning visual performance as a function of age, sex, and other variables. The present paper is one of a series of reports (including Burg, 1966 and 1967a) presenting normative data on vision that should be of interest to others working in this area.

METHOD

Apparatus

The device used for testing lateral visual field was an American Optical Company screening perimeter, fixed in a horizontal position (Figure 1). The perimeter has a 30-cm.-radius black mat arc, with recessed 4-mm. white targets located at 5° intervals around the arc. Each of the targets is activated by depressing a spring-loaded plunger on the back side of the arc, and the target disappears from view as soon as the plunger is released. Each target subtends approximately 46 min. of arc. At each end of the arc is a sighting device for obtaining the correct eye level and 30 cm. distance from the cornea to the fixation point (a 4-mm. white target in the center of the arc). An adjustable double chin rest is used to obtain this proper positioning of *S*'s head for monocular testing. The arc is evenly illuminated by a 50-w., 120-v. incandescent bulb mounted above *S*'s head on the axis of curvature of the arc.[2] The targets are approximately 69 cm. from the center of the light source. A black cloth is suspended from the lower edge of the arc to screen out *E*'s arm movements when he is manipulating the target plungers.

Three testing units were in simultaneous operation during the course of the data-collection phase of

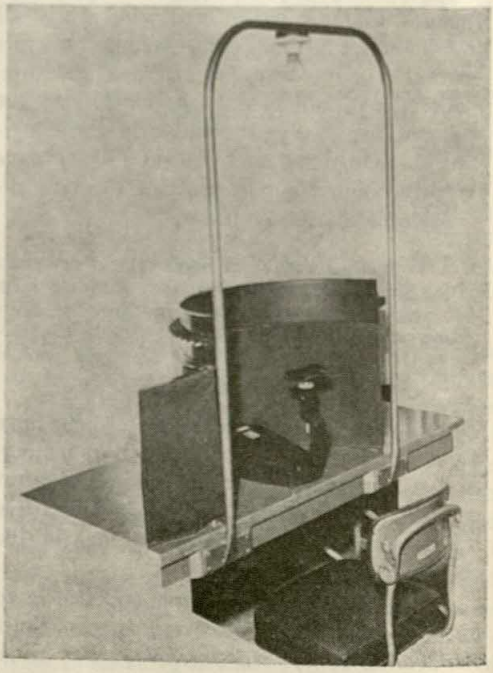


FIG. 1. Lateral visual field screening apparatus.

[1] Data collection and analysis costs were partially borne by the United States Public Health Service (Grant AC-00015), and the State of California and United States Bureau of Public Roads (California Standard Agreement 13600).

[2] Target illumination is 7.5 ftc. measured perpendicular to the light source-to-target axis. When measured perpendicular to the subject-to-target line of sight, target illumination is approximately 2 ftc.

the study. To insure comparable scores, the perimeter test setup was identical in every respect in all situations. All testing was conducted in light-free rooms, with the perimeter arc solely illuminated by the aforementioned 50-w. bulb. (Needless to say, the several test operators employed during the 32-mo. data-collection period were all highly trained to administer the test in an identical manner.)

Procedure

After *S* was seated at the perimeter (on an adjustable-height chair), he rested his chin on the right side of the chin rest (covered by a tissue), placing his left eye in line with the fixation target in the center of the arc. Using the sighting devices at each end of the arc, *E* adjusted the chin rest and the tilt of *S*'s head until the latter's eye was at the proper level and distance from the fixation target.

The *S* then was asked to hold a folded tissue over his right eye. Care was taken that the left-eye nasal field was not intruded upon, either by the tissue or by *S*'s hand. Standing behind the perimeter, *E* asked *S* to fixate with his left eye upon the fixation target (*E* monitored *S*'s adherence to this instruction throughout the test). Using a modified method of limits, *E* then established the (lateral) nasal and temporal limits of *S*'s left-eye field, by recording the position of the furthest target perceived by *S*, both nasally and temporally. If it appeared necessary, measurements were repeated to insure that *S*'s responses were accurate and that his eyeglass frames, if any, were not obscuring the targets.[3] The *S* then placed his chin in the left side of the chin rest, and his right-eye nasal and temporal fields were similarly measured.

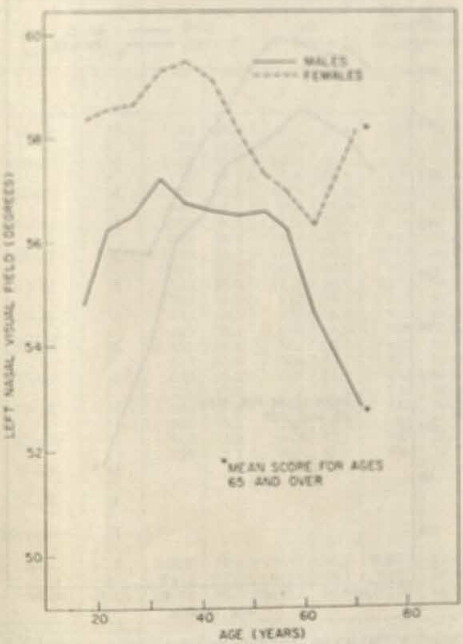


FIG. 2. Left nasal visual field by age and sex.

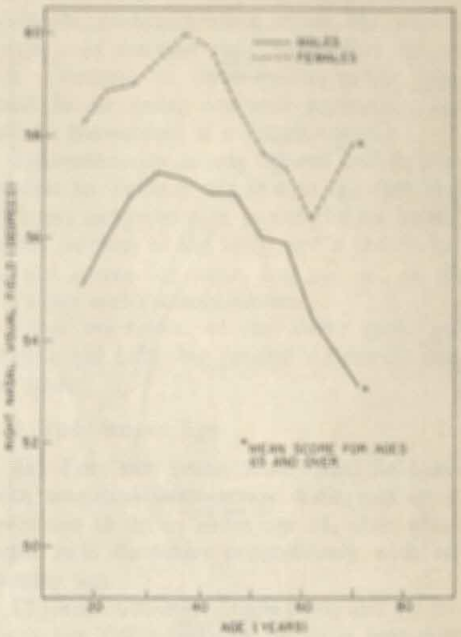


FIG. 3. Right nasal visual field by age and sex.

Subjects

The *Ss* were California drivers who were voluntarily participating in the vision and driving study mentioned earlier. A total of 17,479 *Ss* were involved in the program, 62.8% of whom were male and 37.2% female. The age range was from 16 to 92, and (corrected) static acuity ranged from 20/13 to 20/200. The *Ss* were tested at branch offices of the California Department of Motor Vehicles scattered throughout the State.

RESULTS

Table 1 presents a summary, by age and sex, of mean monocular nasal and temporal fields, as well as total monocular and binocular fields. Figures 2 through 8 present graphical depictions of these data.

The tabular and graphical data are based on somewhat less than 17,300 *Ss*; approximately 200 *Ss* were not included in the present analyses, either because they did not have

[3] The *Ss* were tested with corrected vision, if they indicated that they wore corrective lenses *to drive*.

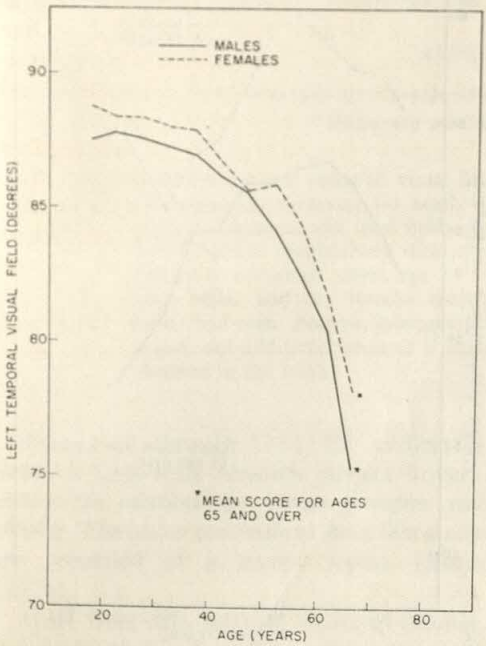


FIG. 4. Left temporal visual field by age and sex.

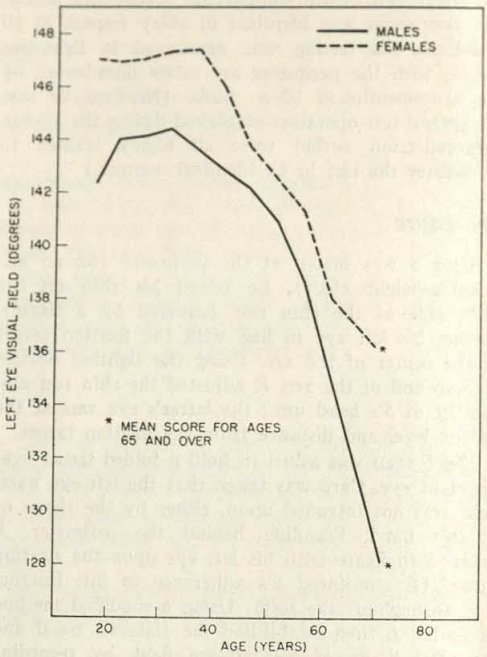


FIG. 6. Left-eye visual field by age and sex.

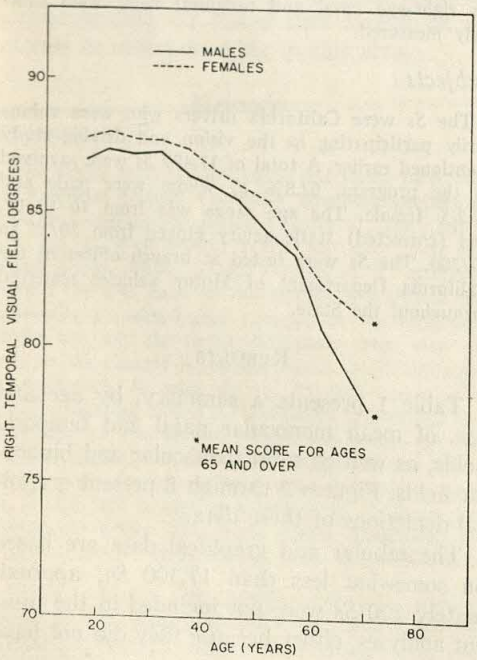


FIG. 5. Right temporal visual field by age and sex.

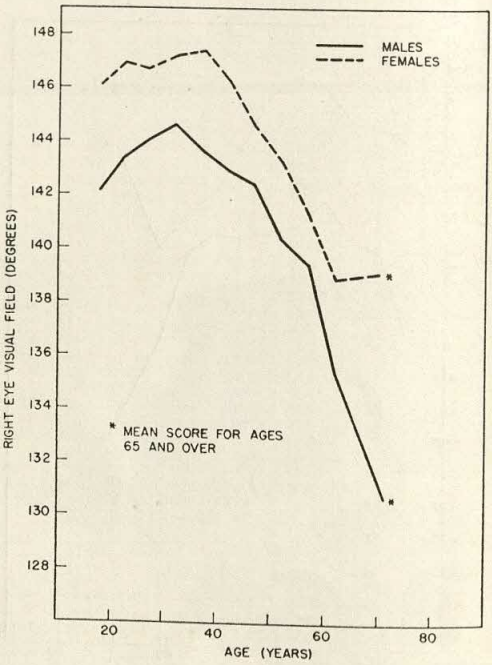


FIG. 7. Right-eye visual field by age and sex.

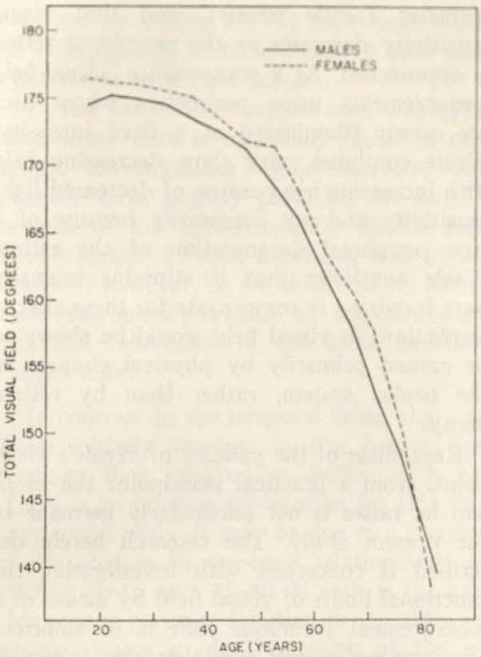


FIG. 8. Total visual field by age and sex.

valid perimeter scores (e.g., they failed to maintain steady fixation upon the central target), or because they were blind in one eye. (Because of their special nature, one-eyed *S*s are being analyzed separately, and will be the subject of a future report.)

Inconsistencies among the cell sample sizes evident in Table 1 are due to the fact that different computer runs provided data for different sections of the table, and a few *S*s had invalid scores for some, but not all, of the variables under consideration.

Upon inspection of the above table and graphs, the following general statements may be made:

1. Field versus Age:

(*a*) For both sexes, total visual field and both temporal fields are at their maximums from age 16 up to about age 35, after which point field decreases progressively with advancing age.

(*b*) For both sexes, nasal fields and (therefore) eye fields increase up to a maximum

TABLE 1

SUMMARY OF MEAN VISUAL FIELD SCORES BY AGE AND SEX

Age	Sex	Left nasal field (in degrees)		Right nasal field (in degrees)		Left temporal field (in degrees)		Right temporal field (in degrees)		Left-eye visual field (in degrees)		Right-eye visual field (in degrees)		Total visual field (in degrees)		
		N	$\bar{X}$	*N*	$\bar{X}$	*N*	$\bar{X}$	*N*	$\bar{X}$	*N*	$\bar{X}$	*N*	$\bar{X}$	*N*	$\bar{X}$	*SD*
16–19	M	1181	54.9	1181	55.1	1181	87.5	1181	87.0	1181	142.4	1181	142.1	1182	174.5	9.1
	F	732	58.3	732	58.3	732	88.7	732	87.8	732	147.0	732	146.1	732	176.4	5.5
20–24	M	1380	56.2	1381	56.0	1380	87.8	1381	87.3	1380	144.0	1381	143.3	1379	175.2	7.7
	F	775	58.5	776	58.9	775	88.4	776	88.0	775	146.9	776	146.9	776	176.1	8.8
25–29	M	1159	56.5	1159	56.8	1159	87.6	1159	87.2	1159	144.1	1159	144.0	1158	174.9	7.5
	F	664	58.7	664	59.0	664	88.3	664	87.7	664	147.0	664	146.7	665	176.0	9.0
30–34	M	1118	57.2	1117	57.3	1118	87.2	1117	87.3	1118	144.4	1117	144.6	1116	174.6	7.7
	F	629	59.3	629	59.5	629	88.0	629	87.7	629	147.3	629	147.2	631	175.6	7.2
35–39	M	1118	56.7	1114	57.2	1118	86.9	1114	86.4	1118	143.6	1114	143.6	1114	173.4	9.8
	F	691	59.5	692	60.0	691	87.9	692	87.4	691	147.4	692	147.4	690	175.1	6.7
40–44	M	1121	56.6	1120	56.9	1121	86.2	1120	86.0	1121	142.8	1120	142.9	1119	172.5	9.2
	F	718	59.2	717	59.7	718	86.7	717	86.6	718	145.9	717	146.3	717	173.5	7.6
45–49	M	937	56.5	937	56.9	937	85.6	937	85.5	937	142.1	937	142.4	935	171.2	8.6
	F	640	58.1	640	58.6	640	85.6	640	86.0	640	143.7	640	144.6	640	171.8	10.0
50–54	M	852	56.6	852	56.0	852	84.2	852	84.3	852	140.8	852	140.3	849	168.4	11.5
	F	577	57.3	576	57.7	577	85.9	576	85.5	577	143.2	576	143.2	576	171.5	9.7
55–59	M	633	56.2	629	55.9	633	82.4	629	83.5	633	138.6	629	139.4	629	165.9	13.4
	F	374	57.0	375	57.4	374	84.3	375	83.9	374	141.3	375	141.3	374	168.0	10.3
60–64	M	488	54.6	488	54.5	488	80.4	488	80.9	488	135.0	488	135.4	487	161.5	16.9
	F	265	56.3	265	56.4	265	81.7	265	82.5	265	138.0	265	138.9	265	163.7	14.4
65–69[a]	M	883	52.8	883	53.1	883	75.3	883	77.6	883	128.1	883	130.7	401	158.0	16.5
	F	350	58.1	350	57.9	350	78.1	350	81.2	350	136.2	350	139.1	205	160.6	12.5
70–74	M													266	152.8	16.3
	F													95	157.6	13.8
75–79	M													136	147.7	21.2
	F													36	150.0	20.1
80+over	M													76	139.5	21.4
	F													13	138.5	21.5
All	M	10870	55.6	10861	55.7	10870	85.3	10861	85.2	10870	140.9	10861	140.9	10849	170.8	12.4
	F	6415	57.8	6416	58.1	6415	86.6	6416	86.2	6415	144.4	6416	144.3	6409	172.9	10.3
All	Both	17285	56.4	17277	56.6	17285	85.8	17277	85.6	17285	142.2	17277	142.2	17249	171.6	11.7

[a] Values given are for ages 65 and over for all measures except *total* visual field.

occurring about age 35–40, followed thereafter by progressive decreases with increasing age, with the exception of an increase for females age 65 and over.

(*c*) For both sexes, right-temporal and right-eye visual fields are smaller than their left counterparts for the younger age groups, while the reverse is true for the older age groups. This effect is not exhibited for the nasal fields.

2. Field versus Sex:

(*a*) With the exception of the (relatively small) over-80 age group, females consistently demonstrate larger nasal and temporal visual fields (and, consequently, larger eye fields and total fields) than do men. The greatest difference is evident in the nasal fields.

(*b*) For both nasal and eye fields, females achieve their maximum values between the ages of 35 and 40, while for males this maximum occurs between ages 30 and 35.

An additional finding is that the product-moment correlation between total visual field and age is slightly but significantly less for females (−.420) than it is for males (−.462).

DISCUSSION [4]

Field versus Age

The results of the present study, showing a progressive decline in visual field with increasing age, are consistent with most previous research studies of a similar nature. However, the *interpretations* of findings of this type have *not* been consistent. As Weale (1963) points out, one reason for this lack of consistency is the fact that senile degeneration of the visual mechanism takes several forms, some of which may intrude upon visual-field measurements and consequently lead to mistaken conclusions as to causal relationships. For example, the assumption usually made states that visual field constricts with age due to peripheral degeneration of the retina. However, it is well known that sensitivity to light diminishes with increasing age, possibly due to the gradual yellowing of the crystalline lens and/or because of reduction in pupil diameter (senile miosis), and that visual sensitivity decreases as the peripheral retina is approached. As a consequence, visual-field measurements using perimeters whose arcs are evenly illuminated at a fixed intensity, Weale continues, may show decreasing field with increasing age because of decreased light sensitivity and *not* necessarily because of a true peripheral degeneration of the retina. Weale concludes that if stimulus intensity were increased to compensate for these effects, restrictions in visual field would be shown to be caused primarily by physical changes in the ocular system, rather than by retinal decay.

Regardless of the validity of Weale's viewpoint, from a practical standpoint the problem he raises is not particularly germane to the present study. The research herein described is concerned with investigating the functional limits of visual field by means of a measurement technique that is of sufficient accuracy *and* ease of administration to permit its use as a high-volume testing device (e.g., for driver-license applicants). Thus, the emphasis is on rapid visual *screening* rather than on detailed clinical evaluation. Practically speaking, if measurement conditions are identical for all *S*s, the fact that visual field constricts with increasing age is of importance *regardless* of the physiological reasons underlying this change. In the performance of a task such as driving, in which visual-field restrictions *do* appear to act as a handicap (Burg, 1967b), establishing the fact of these restrictions is a proper and important function of the licensing agency, while etiological diagnosis and possible remedial measures (if any) are the responsibility of the clinician.

One additional point needs clarification. The visual-field values derived in the present study are of importance primarily as indications of *relative* size as a function of age and sex. It is not intended that the obtained mean scores be considered completely accurate in any absolute sense, for at least two reasons:

1. The screening test used provides scores that may underestimate the true values, because the targets are spaced at 5° intervals. From a statistical standpoint, a closer approximation of the true field values could be

[4] The assistance of Frank A. Brazelton (Optometry) and Glenn O. Dayton (Ophthalmology) in interpretation of the study findings is gratefully acknowledged.

obtained by adding 2.5° to each mean nasal- or temporal-field score, and 5.0° to each mean eye-field or total-field score.

2. The target illumination used was somewhat lower than is commonly found in perimetric investigations and, therefore, may possibly have slightly penalized the older *S* because of the light-sensitivity decline mentioned earlier. If such a penalty was in fact imposed, however, in all likelihood its magnitude was minor, because of the substantial size and reflectance of the white perimeter targets used.

In contrast to the temporal fields, the nasal fields initially increase in size before commencing their progressive decline with age. There is no obvious explanation for this finding; conceivably, it may be due to a physiognomic change in eye-nose relationships occurring up to ages 35–40. If such is the case, this change reverses direction in the older age groups, where gradual recession of the eye (relative enophthalmos) causes the nose increasingly to intrude upon the nasal field (Weale, 1963). Both nasal and temporal fields show marked declines after ages 35–40; if anything, however, the temporal fields exhibit a greater decline (both absolute and relative) than do the nasal fields.

Similarly, at present no readily acceptable explanation can be offered for the reversal in relative size that seems to occur, with age, between the right temporal and monocular fields, and the corresponding fields for the left eye. As yet unpublished monocular visual acuity data for these *S*s do not exhibit corresponding changes with age. As is possible with *any* research study, this finding (as well as other results) may be the consequence of an artifact in the data; however, in view of the large sample size and the extreme care taken to maintain standardized experimental conditions, the likelihood of experimental artifact providing a reasonable explanation for this outcome appears quite small.

Field versus Sex

No obvious explanations can be offered either for the slight but significant superiority in visual-field size demonstrated by females or for the fact that nasal and (consequently) eye fields peak at a later age for females than for males. The fact that visual field correlates with age slightly less for females than for males *can* be explained, however, by the smaller sample size for females, which results in a narrower range of scores (see the standard deviations presented in Table 1) and, consequently, a statistical limitation on the size of the correlation obtainable.

In summary, it may be said that several of the study findings pose interesting questions of interpretation. Additional research obviously is necessary if reasonable answers to these questions are to be found. For example, it would be desirable to obtain field measurements on a large number of *S*s with higher levels of target illumination, physiognomic evaluation, pupil-size measurements, and so on. Data collected in the present study are currently being subjected to more detailed analysis to determine whether, despite the precautions taken, those *S*s who were tested wearing glasses (some 32% of the total) were in fact experiencing restrictions in their field while being measured that could in some way account for some of the findings. Other analyses in the planning stages include the effects of smoking and fatigue (lack of sleep) upon visual performance; however the likelihood is small that either of these analyses will provide satisfactory explanations for any of the findings described above.

REFERENCES

BURG, A. Visual acuity as measured by dynamic and static tests: A comparative evaluation. *Journal of Applied Psychology,* 1966, **50**, 460–466.

BURG, A. Light sensitivity as related to age and sex. *Perceptual and Motor Skills,* 1967, **24**, 1279–1288. (a)

BURG, A. *The relationship between vision test scores and driving record: General findings.* Los Angeles: University of California, Engineering Department, Report 67-24, 1967. (b)

WEALE, R. A. *The aging eye.* London: H. K. Lewis, 1963.

(Early publication received August 21, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 1, 16-21

RATE OF COMPREHENSION OF AN EXISTING TELEPRINTER OUTPUT AND OF POSSIBLE ALTERNATIVES [1]

E. C. POULTON AND C. HELEN BROWN

Applied Psychology Research Unit, Cambridge, England

367 adults were given short passages to study for limited periods of time. They then had to answer open-ended questions on the content. A separate-groups design with groups of about 50 adults and a Greco-Latin-square design using 96 adults appeared to have about equal potential power statistically, but the differences between comparable conditions were almost twice as large with separate groups. Text typed in pica combined upper- and lower-case letters was comprehended on the average about 13% more rapidly ($p < .02$) than text in pica or Siemens all upper-case letters. There was little difference ($p > .05$) between the Siemens all upper case, a dyeline copy of the Siemens, and pica or elite all upper case; and little difference also ($p > .05$) between pica and elite all lower case, which gave intermediate rates of comprehension. The lettering of the question sheets did not affect the results.

Telex and telegram services employ only 26 letter codes, which are printed out by the teleprinter at the receiving end in capitals or upper-case letters. Transmitting messages coded in combined upper- and lower-case letters would be more expensive, since it would require an extra binary unit of information to distinguish between the upper and lower case. However, a printout of combined upper- and lower-case letters might be easier for the receiver to read. This was one of the questions investigated in the two experiments to be reported.

Curiously enough, very little experimental work has been published comparing the typefaces of teleprinters and typewriters, although large amounts of material are being produced in this way and read every day. The only experiments appear to have been by Fox (1963). He compared material typed in standard elite with the same material typed in the smaller-sized capitals of Gothic elite. Gothic elite has two sizes of capitals instead of upper- and lower-case letters. If the smaller size is used for text, and the larger size is reserved exclusively for headings, use of the typewriter shift key is virtually eliminated. This makes typing easier and hence less expensive. In comparing passages of prose typed in the two styles, Fox used as his criterion the speed of reading without measuring comprehension. He found that text typed in small capitals was read 18% more slowly than the standard elite text, but only when read first. On changing to the other style the readers did not change their reading speed appreciably, presumably because they knew that their level of comprehension was not being measured. Similar transfer effects have been reported in other situations (Poulton & Freeman, 1966). Clearly, more work is required in order to assess the size of the likely gain to the reader, if any, of converting teleprinters from all capital letters to combined upper- and lower-case letters.

The classical experiment of Tinker and Paterson (1928) suggests that text typed in combined upper- and lower-case letters ought to be easier for the reader to comprehend than text typed all in capitals. These authors used the Chapman-Cook speed of reading test printed in 10-point "old style" type, presumably without leading. This method also measures speed of reading, but demands a certain level of comprehension since incongruous words in the passage have to be

[1] The authors are grateful to M. A. G. Howgate of the British Government Communications Headquarters for supplying the Siemens and dyeline Siemens teleprinted material used in the experiment. P. R. Freeman kindly supplied the Greco-Latin-square design, and P. M. E. Altham gave advice on how the resulting data were to be analyzed. Financial support from the British Medical Research Council is also gratefully acknowledged.

crossed out. The all-capitals print was found to take 13.4% longer to read. The only other experiment giving statistically reliable results which is in any way relevant is by Paterson and Tinker (1932). They found text typed in American Typewriter elite typeface slower to read than text printed in any of eight printers' typefaces in common use.

A second aim of the experiments was to compare text typed all in capitals with text typed all in lower-case letters. Words in lower case have typical shapes as a result of the ascenders and descenders (Tinker, 1963, p. 60). The shapes might be expected to help a reader even in the absence of capital letters at the start of sentences and for proper names. Apart from tradition, the only reason for using all capital letters rather than all lower-case letters for teleprinters is that capitals are said to perform better when teleprinters are left for prolonged periods without maintenance. However, reasonably large, well-designed, lower-case letters should be equally good in this respect, and might be quicker and easier to read. This point also seemed worth investigating.

A third aim was to compare two different experimental designs, to see which was the more powerful statistically. A separate-groups design avoids transfer effects between experimental conditions like those found by Fox (1963). But its power is reduced by the large sizes of the individual differences if the samples of individuals are at all representative (Poulton, 1965a). In contrast, a Greco-Latin-square design produces unmeasurable transfer effects, which may reduce the sizes of the differences between conditions (Poulton & Freeman, 1966), but large differences between individuals can be eliminated from the statistical comparisons of conditions. Clearly it is useful to know which design is the more powerful, in order to increase the sensitivity of experimental comparisons in this research area (Poulton, 1965b).

A fourth aim was to see whether the style of the question sheets affected the scores for comprehension. In experiments comparing all capitals with mixed upper- and lower-case newspaper headlines, it was found to make no difference to the experimental comparisons whether the question sheets were typed in

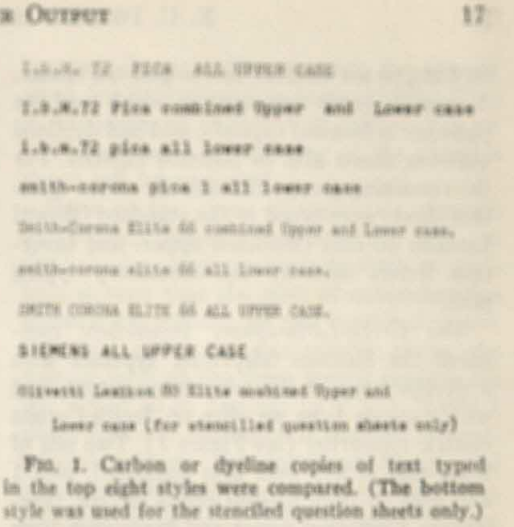


FIG. 1. Carbon or dyeline copies of text typed in the top eight styles were compared. (The bottom style was used for the stencilled question sheets only.)

upper-case or mixed upper- and lower-case letters (Poulton, 1967). However in a pilot experiment (not previously reported) practice pieces of newspaper with all the headlines in capitals were found to favor reliably ($p < .01$) test newspaper sheets with all the headlines in capitals. This suggests that although there may be transfer effects between the styles in which the text is set, the style of type used for the question sheets may make no difference. If so, it could save a lot of effort, especially with Greco-Latin-square designs, since all the question sheets do not then have to be produced in all the styles of type to be compared.

The *separate-groups* experiment compared the existing Siemens teleprinter typeface, which consists of relatively large capitals without serifs (Figure 1), with possible alternative typefaces. To represent the output of a more complex teleprinter system with both upper- and lower-case letters, the pica typeface of the IBM 72 electric typewriter was chosen. To represent the output of a simpler existing system, but with all lower-case letters, the lower-case pica letters of the IBM 72 electric typewriter were used (see Table 1). IBM 72 pica capitals were included as a possible straight alternative to the Siemens capitals. Figure 1 shows that they are serifed, as compared with the Siemens capitals which do not have serifs, but they are not quite so tall. Here the stencils for the question sheets were typed like the text in IBM 72 pica

capitals, to see whether this helped the reader. A separate group was also run which read the passages in Siemens capitals, and had dyelined question sheets also in Siemens capitals. For the remaining groups the stencils of the question sheets were typed in the standard Olivetti Lexikon 80 elite combined upper- and lower-case letters used by the laboratory typing pool.

The *Greco-Latin-square* experiment compared the Siemens teleprinter typeface with a dyeline copy of the teleprinter output, and with capitals from an elite 66 Smith-Corona electric typewriter (see Figure 2). Two sets of lower-case letters were also compared, pica 1 and elite 66 from Smith-Corona electric typewriters. Combined upper and lower case from the Smith-Corona elite 66 electric typewriter was included as a sixth condition to provide a base line. The stencils for all the question sheets were typed in the standard Olivetti Lexikon 80 elite combined upper and lower case.

METHOD

This followed that of previous experiments (Poulton, 1965a). Simple test passages of about 450 words on widespread infectious diseases were typed in the styles shown in Tables 1 and 2 and Figure 1. The same standard teleprinter layout was used throughout: an average line of about 6.5 in. long, triple spacing between lines, no indentations at the beginnings of paragraphs, and about 20 lines to a page. All typing was on four-ply teleprinter paper rolls which have carbon paper interleaved. The first carbon copy on pink paper, the second carbon copy on yellow paper, and occasionally the bottom copy on green paper, were used in the experiment, each person having the same color throughout. The dyeline copies of the Siemens were made from the top sheet which consisted of tracing paper.

There were 10 open-ended questions on each passage to test for comprehension. Normally the questions were typed in upper- and lower-case Olivetti Lexikon 80 elite letters on a wax stencil, and stenciled copies run off. But for the two experimental conditions indicated by Footnote b of Table 1, the lettering of the question sheets corresponded to the lettering of the test passages. The question sheets for the IBM 72 Pica all-upper-case passages were stenciled copies. The question sheets for the Siemens all-upper-case passages were dyeline copies. The questions were spread evenly over the text. There were also passages of about 2,000 words typed in the styles of Table 2. These were used in the Greco-Latin-square experiment to familiarize the readers with the styles before they were given the test passages. The *S*s were tested in groups of about a dozen.

Separate-groups experiment. Each experimental *S* was allocated to one of the styles listed in Table 1. All the passages were typed in all the styles and were read in the same order by all *S*s. After the purpose of the experiment had been explained to them, *S*s were handed two passages each of about 450 words, comparable to those upon which they were to be tested, and told to look them over for 3 min. This was followed by a practice passage which had to be studied for only 75 sec. When time was up, each *S* had to mark with a pencil the point in the passage up to which he had read, and then to answer 10 practice questions on the content of the passage. Unlimited time was available for answering, but *S*s seldom took longer than 4 min. Finally a test passage was read under conditions identical to the practice passage, and 10 test questions on it were answered. The whole experiment took about 25 min. The 18 *S*s who answered correctly 1 or less of the 10 test questions were discarded.

Greco-latin-square experiment. The 96 *S*s read passages typed in each of the six styles listed in Table 2. The six passages, each of which was typed in the six styles, and the six orders of passage-style combinations were arranged in a Greco-Latin

TABLE 1

RESULTS OF SEPARATE-GROUPS EXPERIMENT

Style of type	No. readers	Average comprehension (%)	Average rate of reading (wpm)
IBM 72 pica combined upper and lower case[a]	47	55.1[c]	273
IBM 72 pica all lower case[a]	49	51.9	254
Siemens all upper case[a]	50	49.3	258
Siemens all upper case[b]	52	48.6	265
IBM 72 pica all upper case[b]	73	46.9	262

[a] The question sheets were in the standard Olivetti Lexikon 80 elite upper- and lower-case letters.
[b] The question sheets had the same style of type as the test passages.
[c] Pica combined upper and lower case was reliably better than the three all-upper-case conditions ($p < .02$).

TABLE 2

RESULTS OF GRECO-LATIN-SQUARE EXPERIMENT

Style of type	Average comprehension (%)*
Smith-Corona elite 66 combined upper and lower case	62.9
Smith-Corona pica 1 all lower case	60.5
Smith-Corona elite 66 all lower case	60.3
Siemens all upper case	59.2
Smith-Corona elite 66 all upper case	58.7
Dyeline Siemens all upper case	58.5

* The three upper-case styles were reliably worse than the three remaining styles ($p < .05$).

square. Before each test passage the *S* spent 5 min. looking at a passage of about 2,000 words typed in the same style as the test passage. Ninety seconds were allowed for studying each test passage, and a further 4 min. for answering the 10 questions on it. There was a break of about 15 min. between the first three and the last three tests. The whole experiment took just under 2 hr. to administer. Three *S*s who failed to follow the correct procedure had to be replaced by substitutes.

Experimental subjects. Of the 271 *S*s in the separate-groups experiment, half were men, and of these about one-third were young, enlisted, naval ratings. Overall the ages ranged from 17 to 82 yr. Each of the five experimental groups had about equal proportions of each kind of *S*. Only 8 of the 96 *S*s in the Greco-Latin-square experiment were men. They were allocated randomly to the different experimental conditions. Here the ages ranged from 21 to 59 yr.

RESULTS

Separate groups. The results of the separate-groups experiment are given in Table 1. The upper of the two Siemens groups had stenciled question sheets typed in the standard elite combined upper- and lower-case letters. If the lettering of the question sheets affected the score for comprehension, this is the group most likely to show the effect. For the elite combined upper and lower case of the question sheets is more different from the san-serif Siemens all upper-case text than from any of the other styles of text in the table. The lower of the two Siemens groups in the table (with the superscript b) had dyeline copies of the question sheets produced from the Siemens teleprinter. Here the lettering of the text and question sheets was in the same style. Comparing the two groups, it will be seen that the lettering of the question sheets made little difference ($p > .05$). If anything, the Siemens text with the dyeline Siemens question sheets was comprehended slightly less well than the Siemens text with the stenciled upper- and lower-case question sheets, 48.6% compared to 49.3%.

The pica all-upper-case passages and their stenciled question sheets were made on the same pica electric typewriter in upper-case letters. This may have been even less of a change for the reader than switching from Siemens text to dyeline Siemens questions. The table shows that the pica all-upper-case text was comprehended if anything slightly less well than the Siemens text; the average score for comprehension being 46.9%. Thus in comparing all capitals with combined upper and lower case, the lettering of the question sheets can be neglected.

None of the three all-upper-case versions of the text differed reliably ($p > .05$), so their results can be pooled. Analysis of variance showed that the level of comprehension of the group reading the pica combined upper- and lower-case text was reliably above the average level of the three groups reading the text typed all in capitals ($p < .02$). Table 1 shows that the difference in average comprehension between the combined upper- and lower-case text and the various all-upper-case texts ranged from 5.8 to 8.2%, which become 10.5 to 14.9% when expressed as percentages of the average comprehension of the combined upper- and lower-case text. The level of comprehension of the all-lower-case text lay intermediately and was not reliably different from either. The last column of Table 1 shows that the average rate of reading was a little faster for the combined upper- and lower-case text, but here none of the differences was reliable.

Greco-latin square. The results of this experiment are given in Table 2. Analysis of variance showed that the average level of comprehension of the text typed in the three upper-case styles was reliably below the average level of the text typed in the three remaining styles ($p < .05$). Here Table 2 shows that the difference in average comprehension between the combined upper- and lower-case text and the all-upper-case texts ranged from 3.7 to 4.4%, which become 5.9 to 7.0% when expressed as percentages of the average comprehension of the combined upper- and lower-case text. The all-lower-case text again lay intermediately.

DISCUSSION

The rank orders of the results for comprehension in Tables 1 and 2 are similar. The combined upper- and lower-case text came off best, whether it was typed in pica or elite. The all-upper-case text came off worst, whether the letters were pica or elite and hence serifed, or Siemens or dyeline Siemens without serifs. The all-lower-case text fell between, whether typed in pica or elite. There

was little difference between text typed in the various styles of all-upper-case letters. The pica and elite all-lower-case texts were also about equal. The differences between the combined upper- and lower-case texts and the all-upper-case texts are in line with the results of Tinker and Paterson (1928) referred to in the introduction for passages printed in 10-point "old style."

These results for comprehension suggest that if teleprinters are to have new typefaces, not a great deal is to be gained by changing from the existing all upper-case Siemens letters to the pica or elite all-lower-case letters used in this experiment, as long as the material has to be read and understood. The greatest gain from the point of view of the reader, who has got to understand and take in the information presented, would be a change to combined upper- and lower-case letters. This unfortunately demands a rather more elaborate form of coding for automatic transmission, and hence some additional expense.

Comparing the results of the two different experimental designs, the means for comprehension are rather higher in Table 2 because 90 sec. were allowed for reading each passage instead of 75 sec., and because *S*s had become more practiced at reading and answering the questions by the time they had done so six times over, as required by the Greco-Latin-square design. The differences between the means of the combined upper and lower case and the all upper case in Table 2, 3.7–4.4, are rather smaller than the corresponding differences of 5.8–8.2 in Table 1. This is unlikely to have been due to the use of pica combined upper and lower case in Table 1, and of elite in Table 2. For the pica and elite all lower case gave almost identical results in Table 2. And the small difference between pica all upper case and Siemens in Table 1 in favor of Siemens, corresponds to the small difference between elite all upper case and Siemens in Table 2, again in favor of Siemens. Thus the pica and elite typefaces appear to be about equally effective. The larger differences between conditions in the separate-groups experiment of Table 1, compared with the Greco-Latin-square experiment of Table 2, could have resulted from transfer effects between conditions in the Greco-Latin-square experiment reducing the average sizes of the differences (Poulton & Freeman, 1966).

However, the potential power of the two alternative experimental designs to discriminate statistically between combined upper and lower case and all upper case is probably of about the same order. The Greco-Latin-square experiment did not make full use of the potential power of a factorial design. Its original purpose was to make comparisons between three upper-case alternatives, and between two all-lower-case alternatives. A single combined upper- and lower-case condition was included simply as a base line. If there had been two combined upper- and lower-case conditions, and only two all-upper-case conditions, it would have been possible to have tested directly between three pairs of conditions. A level of confidence at around $p < .02$ might then have been achieved, as in the separate-groups experiment. It took 96 readers for the Greco-Latin square, compared with five groups each of about half the size (a total of 271 readers) in the separate-groups experiment. Increasing the numbers of readers much further becomes a relatively inefficient method of increasing statistical power. It is probably more effective to stratify readers according to reading ability, and to use subgroups of equal size drawn from each of a number of ability ranges. This is the technique used by Zachrisson (1965, Experiment 4) on children starting to read. It would seem to offer the best hope for increasing statistical power without sacrificing the representativeness of the readers.

The relatively small and nonsignificant differences in rates of reading in Table 1 are contrary to the results of Fox (1963) quoted in the introduction. This is probably because our readers knew that they were to be tested for comprehension, while Fox's readers knew that they were not to be tested.

REFERENCES

FOX, J. G. A comparison of Gothic Elite and Standard Elite typefaces. *Ergonomics,* 1963, **6**, 193–198.

PATERSON, D. G., & TINKER, M. A. Studies of typographical factors influencing speed of reading: X Style of type face. *Journal of Applied Psychology,* 1932, **16**, 605–613.

POULTON, E. C. Letter differentiation and rate of comprehension in reading. *Journal of Applied Psychology*, 1965, **49**, 358–362. (a)

POULTON, E. C. On increasing the sensitivity of measures of performance. *Ergonomics*, 1965, **8**, 69–76. (b)

POULTON, E. C. Searching for newspaper headlines printed in capitals or lower-case letters. *Journal of Applied Psychology*, 1967, **51**, 417–425.

POULTON, E. C., & FREEMAN, P. R. Unwanted asymmetrical transfer effects with balanced experimental designs. *Psychological Bulletin*, 1966, **66**, 1–8.

TINKER, M. A. *Legibility of print*. Ames: Iowa State University Press, 1963.

TINKER, M. A., & PATERSON, D. G. Influence of type form on speed of reading. *Journal of Applied Psychology*, 1928, **12**, 359–368.

ZACHRISSON, B. *Legibility of printed text*. Uppsala, Sweden: Almqvist & Wiksell, 1965.

(Received January 23, 1967)

BUREAU OF EDUCATIONAL AND PSYCHOLOGICAL RESEARCH
GOVT. OF WEST BENGAL
LIBRARY
CALCUTTA

Bureau of Edn. & Psyl. Research
(S. C. E. R. T.)
Date 6.3.81
Acc. J 697

Journal of Applied Psychology
1968, Vol. 52, No. 1, 22–28

ORGANIZATIONAL CORRELATES OF EXECUTIVE ROLES [1]

RAYMOND A. KATZELL, RICHARD S. BARRETT, DONALD H. VANN, AND JOHN M. HOGAN [2]

New York University

Relationships between each of 9 executive role dimensions and various features of the organizational setting were studied. The sample consisted of 194 middle-management personnel of the U. S. Army. The roles, as described on questionnaires by the executives themselves, were found to vary with differences in the work setting. Organizational features most prominently associated with variations in role included the mission and level of the organization, the job family of the executive, and his span of control. The role dimensions which were most prone to vary with the situation were staffing, controlling, and time spent with others. Executives whose roles emphasized administrative more than technical features received higher average performance ratings from their superiors.

It is a well-known fact of social psychology that the behavior of members of a social system is shaped by characteristics of the system as well as of the individuals. The concept frequently invoked to link the system and the individual is that of role, that is, a set of activities and responsibilities associated with a particular position or job.

The notion that a given job role may vary in different organizations has important implications for personnel management. Among them is the prospect that the personnel requirements for a job are likely to depend on contextual considerations. Failure to pay due attention to situational factors could therefore result in defective personnel procedures, such as invalid selection techniques, ineffective training programs, or inappropriate supervision.

While the applicability of these concepts to executive personnel management is general recognized, they have received relatively little systematic or detailed attention. It would seem important not only to know in what ways executive roles may differ among organizational settings, but also whether there are stable relationships between given role dimensions and specific organizational variables such as size or span of control. The present paper reports research on this topic, undertaken as an early stage of a program of improved personnel management of civilian executives in the Department of the Army.

More specifically, the study sought to determine whether differences in the ways in which executives perceive their roles are associated with differences in variables characterizing their organizational settings. As subsidiary objectives, the study also sought evidence of interaction between organizational and personal variables in their relations with these role dimensions, and sought data bearing on the relationships between role behaviors and executive effectiveness.

METHOD

Subjects

The data were collected on 194 male civilian executives of the Department of the Army. With the exception of seven who were in General Service (GS) Grades 16 and 17, they were classified in GS Grades 13 through 15, roughly equivalent to "middle management." To assure a sample broadly representative of such positions, *Ss* were selected from six Army organizations with different missions, levels, sizes, and geographical locations. The *Ss*' mean age was 46.7 yr. ($\sigma = 8.1$), and had a mean supervisory experience of 15.2 yr. ($\sigma = 8.6$). The jobs covered a variety of titles in such fields as finance, research and development, technical specialties, and personnel management; relatively few could be considered typically "line management," since such positions are usually occupied by military officers.

[1] This research was conducted by contract to the Department of the Army under sponsorship of the United States Army Personnel Research Office.

[2] The opinions expressed in this paper are those of the authors and do not necessarily reflect official Department of the Army policy.

Measures of Job Behavior

This paper is limited to the study of job behaviors and responsibilities as perceived and reported by *Ss* themselves, using the following questionnaires:

(*a*) Executive Position Description: an adaptation of 85 items from the Executive Position Description: Form IV developed by Hemphill (1960). Respondents indicated on a 4-point scale the extent to which an activity or responsibility described by an item was a part of their position. The following items are typical:

"Schedule work so that it flows evenly and steadily."
"Reduction of costs."

(*b*) Work Analysis Form: an adaptation of one by the same name described by Stogdill and Shartle (1955). The respondent estimated the percentage of time which he spends in contact with others and in individual effort. He then reported the detailed breakdown of the time spent in these two major categories of duties, and also the time spent on duties included in a list of major responsibilities. The following items are typical of each of the categories, respectively:

"Attending committee meetings and conferences."
"Preparing and writing reports, orders, etc."
"Professional consultation—giving technical advice."

(*c*) Performance Style Questionnaire: 40 items listing the work patterns, habits, and interpersonal relationships a person may emphasize in his job following the approach described by Barrett (1961). The respondent rated the items in a forced-normal distribution according to the extent to which they characterize his job behavior. Typical items are:

"Stays out of arguments."
"Gives attention to matters immediately as they arise."

Measures of Organizational Characteristics

Information about the organization as a whole, such as the number of personnel or type of mission, was obtained by the researchers from records and interviews with top-level executives. Descriptions of certain organizational variables immediate to each *S*'s job, such as his GS level, were obtained from him by use of a background information sheet. A third type of contextual information concerned the supervisory climate in which each *S* works. Some of this was obtained by questionnaire from the supervisor himself, including such items as his age, education, amount of supervisory experience, etc. To provide data on the supervisor's behavior, each *S* filled out the Leader Behavior Description Questionnaire (Stogdill & Coons, 1957) describing his perceptions of the extent to which his supervisor manifests consideration and initiation of structure.

Other Measures

One of the subsidiary objectives of the study was to see whether personal characteristics interact with organizational characteristics in their relations with job behavior. For this purpose, data on certain personal variables, such as age and education, were obtained on the background information sheet filled out by each *S*. Another subsidiary objective concerned possible relationships between executives' job behavior and ratings of their effectiveness. For the latter purpose, a simple three-scale rating form was filled out by the immediate supervisor of each *S*. One of the scales concerned a comparison of the job performance of *S* with others at his level, a second entailed a comparison of his performance with standards, and a third asked for the supervisor's willingness to hire the man if he had his choice. This instrument was used in another study of executives by one of the authors, in which the sum of the three scales was found to have an interrater reliability of .77.

Data Analysis

In order to condense the descriptions of job behaviors and responsibilities into a limited number of meaningful role dimensions, a centroid factor analysis was performed on the three questionnaires describing job activities. Excluded were those items in the Executive Position Description where more than 50% of the respondents reported that the activity or responsibility was not relevant to their position. This yielded a total of 90 items which entered into the intercorrelation matrix. Eight orthogonal factors emerged:

1. *Long-Range Planning:* planning for future activity, including forecasting trends or events, conducting pilot projects, being concerned with legislation which might affect the organization, and establishing long-range objectives for the organization.
2. *Staffing:* personnel activity characterized by involvement in selection, placement, training, and advancement of employees.
3. *Technical Consultation:* gathering, consolidating, interpreting, and providing others with facts and information.
4. *Budgeting:* preparation and defense of budgets, and accounting for expenditures.
5. *Shared versus Individual Responsibility:* the shared pole being characterized by emphasis on keeping others informed of present states of affairs, whereas the individual pole emphasizes personal responsibility for taking action.
6. *Operational versus Professional Concerns:* the operational pole being characterized by concern for the efficiency of present organizational operations and personal professional development, whereas the professional pole is oriented toward advising others on technical matters.
7. *Technical versus Administrative Activity:* the technical pole being characterized by the use of professional tools and techniques, including writing

TABLE 1

ROLE DIMENSIONS AS FUNCTIONS OF ORGANIZATIONAL VARIABLES

Organizational variable	N		Role dimension								
			Long-range planning	Staffing	Technical consultation	Budgeting	Shared vs. individual effort	Operational vs. professional	Technical vs. administrative activity	Controlling	Time spent with others
Organization mission											
Administration	63	*M*	10.2	10.6	6.8	6.0	1.6	.8	−.2	4.9	55.6
		σ	5.2	3.8	2.1	4.1	1.5	1.4	1.9	1.6	20.3
Manufacturing, engineering, & material handling	20	*M*	14.0	10.6	7.5	7.8	1.9	1.5	−.3	4.6	64.0
		σ	3.8	4.5	2.0	3.8	1.3	2.3	1.4	1.7	16.2
Research & development	63	*M*	13.0	8.0	6.9	5.3	1.6	.8	−.7	4.0	50.8
		σ	5.0	4.1	2.4	4.2	1.6	2.1	1.7	2.0	21.2
		F	6.96**	6.90**	.61	2.84	.26	1.12	1.14	4.30*	3.32*
Organization level											
Office of Chief of Staff	38	*M*	12.5	7.3	6.7	5.2	1.6	.6	−.9	3.3	53.4
		σ	4.7	4.1	2.5	4.3	1.5	2.2	1.6	1.8	21.6
Command headquarters	27	*M*	13.3	9.4	6.6	5.6	1.3	1.2	−.6	4.8	46.1
		σ	5.0	3.8	2.5	3.7	1.7	1.8	1.8	1.9	18.9
Field installations	81	*M*	11.2	10.6	7.2	6.4	1.8	1.0	−.1	4.9	58.1
		σ	5.3	4.0	2.0	4.2	1.4	1.7	1.8	1.6	19.8
		F	1.92	8.38**	1.01	1.05	1.15	.83	2.53	10.86**	3.65*
Urban-Rural location											
Urban	47	*M*	12.0	7.4	6.6	4.9	1.5	.6	−.8	3.7	51.9
		σ	4.8	4.1	2.6	4.0	1.4	2.2	1.5	2.0	21.2
Rural	99	*M*	11.9	10.5	7.1	6.4	1.7	1.1	−.2	4.8	56.0
		σ	5.3	4.0	2.1	4.2	1.6	1.7	1.8	1.7	20.2
		t	.17	4.42**	1.10	1.98*	.50	1.34	1.99*	3.56**	1.15
Washington-Other location											
Washington	38	*M*	12.5	7.3	6.7	5.2	1.6	.6	−.9	3.3	53.4
		σ	4.7	4.1	2.5	4.3	1.5	2.2	1.6	1.8	21.6
Other	108	*M*	11.8	10.3	7.0	6.2	1.6	1.0	−.3	4.8	55.1
		σ	5.3	4.0	2.2	4.1	1.5	1.7	1.8	1.7	20.3
		t	.75	3.88**	.71	1.23	.21	1.20	1.98*	4.67**	.45
Organization size											
Larger	121	*M*	11.6	9.6	6.9	6.0	1.6	.9	−.4	4.3	56.3
		σ	5.0	4.3	2.3	4.2	1.5	1.9	1.7	1.8	20.4
Smaller	25	*M*	13.8	9.2	7.2	5.4	1.6	1.1	−.3	4.9	46.8
		σ	5.2	3.8	2.2	4.0	1.7	1.9	1.7	2.0	19.9
		t	2.01*	.38	.63	.74	.23	.59	.31	1.44	2.11*
Change in size											
Expanding	99	*M*	11.9	10.5	7.1	6.4	1.7	1.1	−.2	4.8	56.0
		σ	5.3	4.0	2.1	4.2	1.6	1.7	1.8	1.7	20.2
Stable	9	*M*	10.2	7.7	6.3	3.8	1.3	.7	−.6	5.2	45.6
		σ	4.7	3.8	2.7	2.3	1.1	1.9	1.4	2.0	18.2
Contracting	38	*M*	12.5	7.3	6.7	5.2	1.6	.6	−.9	3.3	53.4
		σ	4.7	4.1	2.5	4.3	1.5	2.2	1.6	1.8	21.6
		F	.71	9.75**	.70	2.39	.22	.89	2.11	11.12**	1.16

Table 1—*Continued*

Organizational variable	N		Role dimension: Long-range planning	Staffing	Technical consultation	Budgeting	Shared vs. individual effort	Operational vs. professional	Technical vs. administrative activity	Controlling	Time spent with others
Type of executive job											
Research & development	51	*M*	9.8	10.4	6.7	5.6	1.8	1.3	−.5	4.5	59.0
		σ	4.5	4.1	2.2	4.0	1.4	1.4	1.3	1.7	18.4
Construction & equipment engineering	17	*M*	10.6	9.3	7.4	7.3	2.6	.0	.1	5.2	54.4
		σ	5.8	3.7	2.7	4.0	1.2	1.6	2.5	1.7	18.8
Technical specialties	25	*M*	13.4	9.1	6.6	6.1	1.6	2.0	−.8	4.8	56.2
		σ	3.9	4.3	1.8	4.2	1.7	2.0	1.6	1.8	20.0
Personnel	26	*M*	15.1	8.7	7.5	3.6	1.5	1.3	−.3	4.2	50.1
		σ	4.4	4.4	2.2	3.3	1.3	1.9	2.0	1.7	23.6
Finance & data processing	22	*M*	13.1	9.5	6.4	8.7	.9	1.8	−.3	4.0	48.0
		σ	5.0	4.4	2.2	3.7	1.3	1.3	1.6	1.8	22.6
		F	6.8**	.78	1.10	5.53**	3.83**	8.05**	.71	1.57	1.47

* $p < .05$.
** $p < .01$.

technical reports, whereas the administrative pole involves coordination of activity, communication, and recommendations.

8. *Controlling:* relating to formulation of limited plans or specifications in regard to resources and finances of the organization, including reducing costs.

Only a few of the items from the Performance Style Questionnaire were found to have appreciable loadings on any of these factors, so for reasons of efficiency it was decided to exclude it from the measurement of job behavior. Scores on the eight factors were computed for each *S* on the basis of specification equations, each including between 5 and 10 items from the other two questionnaires. The percentage of time on the job that the incumbent reported spending in contact with other individuals was considered as an additional role variable, because of its intrinsic interest. Thus, a total of nine variables constituted executive "job behavior" or "roles" as discussed below.

In passing, it may be of interest to compare the 8 factors described above with the 10 which Hemphill (1960) identified in his study of executives. Long-Range Planning is essentially duplicated in both studies, and there is some similarity between our Controlling factor and his of Business Control. The other factors are not parallel in the two studies, perhaps because of differences in the jobs, in *S*s, in the settings, or some combination of such parameters.

Analysis of variance was used to determine whether each of the nine role dimensions was differentially associated with organizational characteristics. In instances where organizational variables yielded continuous distributions, their association with the nine role dimensions was determined instead by product-moment coefficients of correlation. The *N* on which these analyses were based was usually somewhat smaller than the total of 194, because data were not complete for all *S*s on all variables; no systematic basis was discerned for the loss of cases.

RESULTS

Role Dimensions in Relation to Organizational Variables

Table 1 presents, in terms of analysis of variance, the comparison of role-dimension scores associated with various organizational and job settings. Table 2 reports product-moment correlation coefficients between role-dimension scores and certain continuously distributed organizational variables. In both tables, significance levels are reported on the basis of two-tailed tests.

Tables 1 and 2 show that there were a number of significant differences in roles of the executives studied, depending on the

organizational setting in which they operated. Organizational features associated with these differences included the organizational mission, level, size, stability, and whether urban or rural in location. The executive's job family, GS level, and span of control also were influential. The role dimensions most frequently varying with organizational characteristics were Staffing, Controlling, and Time Spent with Others. These relationships are described in more detail below.

As shown in Table 1, the roles of those executives who worked in an organization having a mission of administration entailed relatively little emphasis on Long-Range Planning and relatively strong emphasis on Staffing and on Controlling. Roles in organizations having a major mission in the area of manufacturing, engineering, or materials handling, tended to be relatively high on Long-Range Planning, on Staffing, and on Time Spent with Others, and were average on Controlling. Those in the laboratory and research functions were moderately high on Long-Range Planning and low on Staffing, Controlling, and Time Spent with Others.

The level of the organization was also related to the roles reported by the incumbents. The higher the level, the less prominent were Staffing and Controlling dimensions.

Those executives who worked in the Washington headquarters gave less emphasis to Staffing, Technical Activity, and Controlling. These results are similar to those comparing urban and rural locations, partially because Washington was one of the major urban settings studied.

Executives in the smaller organizations placed more emphasis on Long-Range Planning, possibly because small organizations may receive less attention and direction from headquarters and may therefore be more autonomous with respect to plans. These executives also spent less time with others, which may support the common observation that larger organizations require more communication. Not surprisingly, executive roles in expanding organizations were relatively high in Staffing responsibilities.

The job family of the executive's position was found to be related to his role: those in research and development work were less involved in Long-Range Planning, those in construction and equipment engineering concerned more with Budgeting and less with Operational considerations, personnel specialists more concerned with Long-Range Planning but less with Budgeting, and financial people more concerned with Long-Range Planning and Budgeting.

As shown in Table 2, the more people supervised directly or indirectly by the executive, the more role emphasis there was on Long-Range Planning, Staffing, Budgeting, Controlling, Administration, and Time with Others.

The GS levels covered in this study were largely limited to Grades 13, 14, and 15. This restriction of range accounts in part for the relatively small correlations found in Table 2 between GS level and role characteristics. Nevertheless, the higher-level people were found to place significantly less stress on Controlling and more on Administrative activity rather than Technical activity or

TABLE 2

COEFFICIENTS OF CORRELATION BETWEEN ROLE DIMENSIONS AND CERTAIN ORGANIZATIONAL VARIABLES

Organizational variable	Role dimension								
	Long-range planning	Staffing	Technical consultation	Budgeting	Shared vs. individual effort	Operational vs. professional	Technical vs. administrative activity	Controlling	Time spent with others
No. persons reporting directly	.08	.60**	−.02	.17*	.07	.10	.01	.19	.32**
No. persons reporting directly or indirectly	.31**	.32**	−.08	.24**	.06	.10	−.20*	.22*	.18*
GS level	.15	.11	−.16*	.13	−.03	−.11	−.29**	−.22*	.16*

Note.—$N = 145$.
* $p < .05$.
** $p < .01$.

TABLE 3

COEFFICIENTS OF CORRELATION BETWEEN ROLE DIMENSIONS AND SUPERVISORS' CHARACTERISTICS

Supervisors' characteristic	Role dimension								
	Long-range planning	Staffing	Technical consultation	Budgeting	Shared vs. individual effort	Operational vs. professional	Technical vs. administrative activity	Controlling	Time spent with others
Initiation of structure	−.17*	.05	−.15	−.01	−.15	−.08	.08	−.01	.14
Consideration	.02	.03	−.05	.07	−.26**	.06	−.07	−.03	.00
Percentage of *S*'s duties supervisor can perform	.07	.04	−.02	.14	.09	.30**	−.12	−.06	−.03
Time *S* served under supervisor	−.22**	.04	.03	−.02	.02	−.01	.06	.08	.06
Time in present position	−.19*	.06	−.01	−.06	−.12	−.05	.08	.00	−.02
Age	−.07	−.01	−.08	−.11	−.08	−.07	−.09	.04	−.07
Amount of education	−.17*	−.05	−.14	−.03	−.12	−.19*	.04	−.26**	.07
Academic standing	.13	−.11	.07	.04	−.03	.07	.01	.10	−.04
Closeness of contact with *S*	.15	−.06	.00	.09	.06	.11	−.14	−.01	.03
Length of face-to-face contact	−.21**	.05	.00	.05	.10	−.09	.16**	.12	.08

Note.—N = 136–146.
* $p < .05$.
** $p < .01$.

Consultation. There was some trend also for them to engage in more Long-Range Planning and to spend more Time with Others.

Table 3 reports the findings bearing on the supervisory aspect of the organizational climate within which *Ss* operated. The table lists correlation coefficients between characteristics of *Ss*' supervisors and *Ss*' own role perceptions. The coefficients in Table 3 are small, and the number of significant ones is not much beyond chance expectancy. There is some indication that the supervisors' formal education was appreciably associated with the roles of their executive subordinates, the more highly educated supervisors having subordinates whose roles involved less emphasis on Long-Range Planning and on Controlling, and greater emphasis on Professional rather than Operational Concerns. This finding may be as much a function of the technical nature of the jobs held by subordinates of highly educated supervisors as it is of the personal impact of the supervisor. On comparing the columns of Table 3, it appears that Long-Range Planning is the role dimension most frequently associated with characteristics of the supervisor; executives whose roles more heavily featured this characteristic were those whose supervisors did less Initiation of Structure, had less extensive face-to-face contact with their subordinates, showed evidence of greater mobility in jobs, and had less education.

Interaction Effects of Personal and Organizational Variables

In a complete nomology of executive role behavior, one would expect that organizational influences would interact with personal dispositions. The present study was not designed to delve deeply into personal correlates of executive roles, but it did include the collection of certain biographical data which could be examined in this light. The personal variables found to be more strongly associated with role have to do with educational background. For example, *Ss* with more formal education described their jobs as entailing less Long-Range Planning, Staffing, Budgeting, and Controlling, but more Professional Concerns. However, most of the personal variables were found to have only slight relationships, if any, with the nine role dimensions; this statement covers variables like age, amount of supervisory experience, and number of dependents. Perhaps for this reason, when two-way analyses of variance were performed of role-dimension scores against both personal and organizational variables, the number of significant interaction effects was only about chance, and yielded no meaningful overall picture.

Relation of Role to Performance Ratings

One of the nine role dimensions turned out to have low but statistically significant relationships to the three ratings (median $r = .19$), to wit, executives whose roles more prominently featured Administrative as contrasted with Technical Activity received higher ratings from their superiors. This was not due to a tendency to rate higher those who occupy higher-level jobs, since the correlations of ratings and GS level were approximately zero. It may signify that administrative behavior is perceived as a more critical requirement of executive jobs of the types studied.

DISCUSSION

In line with psychological theory, executive role was found to be associated with features of the organizational and work setting. In this sample of positions and organizations, differences in executive roles were especially marked between organizations of different missions and levels, between jobs designated as general administration compared to those characterized as scientific or fiscal, and between positions showing differences in span of control. The role dimensions of Staffing, Controlling, and Time with Others were particularly prone to vary with organizational characteristics.

These results underscore the potential fruitfulness of taking organizational settings more fully into account when designing executive personnel management practices, such as techniques of selection, training, or supervision. For example, it would be reasonable to expect that critical requirements for executives having large spans of control, which were found here to be characterized by greater role emphasis on administrative, planning, and personnel matters, would be different from those having smaller spans of control. The logical next step would be to see whether improvements in executive personnel management can indeed be effected when organizational variables are included as moderating parameters in investigations of the relationships between management methods and criteria of effectiveness. The present study indicates the relative potency of a number of organizational variables in this regard; by specifying how role behavior is associated with the more influential variables, it suggests which aspects of management method may be fruitfully examined for their impact.

REFERENCES

BARRETT, R. S. The agreement scale: A preliminary report. *Personnel Psychology,* 1961, 14, 151–165.

HEMPHILL, J. K. *Dimensions of executive positions.* Columbus, Ohio: Bureau of Business Research, Ohio State University, 1960.

STOGDILL, R. M., & COONS, A. E. (Eds.) *Leader behavior: Its description and measurement.* Columbus, Ohio: Bureau of Business Research, Ohio State University, 1957.

STOGDILL, R. M., & SHARTLE, C. L. *Methods in the study of administrative leadership.* Columbus, Ohio: Bureau of Business Research, Ohio State University, 1955.

(Received January 26, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 1, 29-30

USE OF HIGH SCHOOL DATA TO PREDICT WORK PERFORMANCE[1]

MARSHALL H. BRENNER

Lockheed-California Company

For a sample of high school graduates employed in an aircraft plant, teachers' work habits and cooperation ratings, absenteeism, and grade-point data were obtained from high school transcripts and related to work-performance criteria of supervisory ratings, absenteeism, and tardiness records. Significant relationships were obtained between the high school predictors and the work-performance criteria.

For the great majority of industrial test-validation studies, the obtained validities are disappointingly low, indicating that the industrial researcher must search for additional, nontest predictors to be used in conjunction with the test information. One of the factors that is frequently considered important is that of "attitude toward work." Attitude is not frequently utilized, though, because psychologists have found it extremely difficult to obtain useful measures of attitude in an employment setting. However, a meaningful measure of attitude might be obtained from an earlier situation similar in nature to the work situation, one in which the *S* had an opportunity to express similar attitudes. This study was directed toward the investigation of one source of previous evidence of attitude—data concerning behavior in the high school.

Procedure

The *S*s used in this study were 100 employees of the Lockheed-California Company who (*a*) graduated from a high school in the Los Angeles City School Districts, (*b*) attended that school for at least 2 yr., (*c*) were employed by the Lockheed-California Company within 1 yr. of graduation, and (*d*) were employed by the Lockheed-California Company for at least 1 yr.

The data for the first seven variables (the predictors) in Table 1 were obtained from *S*s' high school transcripts, while the data for the next five (the criteria) were gathered from supervisory ratings and departmental records.

The teachers' work habits and cooperation ratings were given on a 3-point scale, and the supervisory ratings on a 5-point scale. The "sex" variable was scored 0 for females and 1 for males.

Results

The means, the standard deviations, and the intercorrelation matrix are presented in Table 1.

Reference to Table 1 reveals the following:

1. The standard deviations of both the predictor and the criterion variables are large enough to provide adequate discrimination.

2. The standard deviations of the absenteeism and tardiness data are much larger than would be expected if the distributions were normal. This indicates that a relatively large percentage of *S*s had either a very low number of absences or a very large number. Plotting of the data revealed this to be the case.

3. The teachers' ratings are very highly related to each other and to students' grade point.

4. The teachers' ratings, the students' grade point, and school absences are significantly related to supervisory ratings. Ten of these 12 correlations are significant at the .01 level.

5. The teachers' ratings and school absences are related to work absences at the .01 level.

6. Grade point and rank in class are extremely highly correlated, indicating that a given grade-point average would be in the same relative rank position in each of the 12 schools furnishing data.

[1] Extensive assistance was received in connection with this study from many members of the Los Angeles City School Districts. It was a pleasure to work with Muriel Sheldon, Clyde Sorensen, and Rose O'Connor of the Secondary Guidance and Counseling Section who provided invaluable advice on the design of the study and obtained all of the predictor data used in it.

TABLE 1

MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS OF VARIABLES

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. School absences	21.5	19.6		40	−11	−14	−28	−27	−32	−25	−31	−21	30	20	06
2. Unexcused school absences	1.8	3.3	40		−04	−06	−18	−16	−25	−11	−31	−09	13	07	17
3. Three-year grade point	2.2	.6	−11	−04		96	78	67	73	36	37	34	−16	10	−58
4. Rank in class (inverted)	42.7	27.4	−14	−06	96		73	63	72	33	34	30	−16	10	−56
5. Teacher's work habits rating	2.1	.4	−28	−18	78	73		85	75	34	44	41	−26	−02	−51
6. Teacher's cooperation rating	2.4	.3	−27	−16	67	63	85		64	30	45	39	−29	−02	−42
7. Grade point last two semesters	2.4	.6	−32	−25	73	72	75	64		30	38	26	−16	07	−41
8. Supervisor's ability rating	3.4	.9	−25	−11	36	33	34	30	30		61	72	−14	−05	−23
9. Supervisor's conduct rating	3.5	.9	−31	−31	37	34	44	45	38	61		73	−31	−15	−22
10. Supervisor's production rating	3.4	.9	−21	−09	34	30	41	39	26	72	73		−23	−14	−29
11. On-the-job absences	11.2	9.2	30	13	−16	−16	−26	−29	−16	−14	−31	−23		37	−03
12. On-the-job tardies	3.6	4.1	20	07	10	10	−02	−02	07	−05	−15	−14	37		−22
13. Sex			06	17	−58	−56	−51	−42	−41	−23	−22	−29	−03	−22	

Note.—A correlation of .20 is necessary for significance at the .05 level. A correlation of .26 is necessary for significance at the .01 level. Decimals have been omitted from the correlation part of the table.

DISCUSSION

These results indicate the following:

1. The high school data collected in this study have adequate discrimination to serve as predictor data. The teachers are making meaningful discriminations of work habits and cooperation behavior.

2. The high school data are significantly related to performance on the job as evaluated by the immediate supervisor. Of even greater significance, the degree of relationship between the high school data, especially the teachers' ratings, and the work-performance data is greater than that normally obtained with aptitude tests (e.g., see Ghiselli & Brown, 1955, p. 227). In other words, whatever it is that determines a supervisor's evaluation of work performance, the high school teacher is able to (and does) measure it more successfully than aptitude tests.

3. Absenteeism on the job can be predicted to a statistically significant extent by absenteeism at school and by the teachers' ratings.

REFERENCE

GHISELLI, E. E., & BROWN, C. W. *Personnel and industrial psychology*. New York: McGraw-Hill, 1955.

(Received February 6, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 1, 31–35

VOCATIONAL INTERESTS OF FEMALE COMPUTER PROGRAMMERS

DALLIS K. PERRY AND WILLIAM M. CANNON

System Development Corporation, Santa Monica, California

Scores of 293 female computer programmers on the SVIB for men were found to be quite similar to those of male programmers, although interests of the women were somewhat higher in aesthetic and scientific fields and lower in technical and technical-supervision occupations. Female programmers indicate interest in all forms of mathematics and lack of interest in people. A key composed of items with a 12% or larger difference between female programmers and other professional women showed a 20% overlapping of professional women in general with a cross-validation group of female programmers. Programmers who are dissatisfied with their occupation score significantly lower on the key than do satisfied programmers.

In the process of measuring the vocational interests of computer programmers and developing a Strong Vocational Interest Blank (SVIB) scale to represent these interests, a substantial number of completed SVIBs for men were obtained from female programmers. The following points became clear through analysis of SVIB scores for the female programmers:

(*a*) The vocational interests of female programmers are highly similar to those of male programmers.

(*b*) The differences in interests that characterize men and women in general characterize male and female programmers as well, with the result that the male programmer key with male norms does not adequately represent the interests of female programmers (Perry & Cannon, 1967).

Because there is considerable overlapping of items between the SVIBs for men and for women, it appeared feasible to use the responses of female programmers on the SVIB for men to develop a computer programmer key for women that could be used with the women's form of the blank. At the same time it seemed desirable to provide normative data to facilitate the interpretation of women's scores on the male programmer key in those instances in which the men's SVIB form is used with women. The present study, therefore, was intended to (*a*) develop a computer programmer key for the SVIB for women, and (*b*) develop female norms for the programmer key for the SVIB for men.

METHOD

Subjects. The 1,378 computer programmers participating in the original study included 186 females. This group was augmented by 107 additional females from six of the large programming organizations in the original study. The same data-collection and reporting procedures were again followed to assure anonymity of participants (Perry & Cannon, 1967). The original data were gathered in 1964 and the additional data in 1966.

The total group of 293 female programmers was reduced for purposes of key construction and validation by dropping 31 members who indicated that they (*a*) were dissatisfied with programming ("probably" or "definitely" would prefer another field); (*b*) were primarily responsible for computer operation, programmer instruction, or computer installation management; or (*c*) had less than 1 yr. of experience in programming. With the exception of the last restriction, these constraints are the same as those applied to the male programmers. Male programmers with less than 2-yr. experience were found to have vocational interests nearly identical with those of more experienced programmers. Consequently, it appeared justifiable to lower the experience requirement from 2 yr. to 1 yr. in the interest of maintaining sample size. The resulting primary sample of 262 female programmers was considered adequate for key construction and validation, although it was smaller than desired.

Sample characteristics. The *S*s in the primary sample ranged in age from 21 to 59, with a median of 29.9, mean of 31.6, and standard deviation of 7.3. Almost 90% were 40 or younger. This group is thus younger than all the present SVIB female criterion groups except secretary-stenographer (Strong & Campbell, 1966). Like male programmers, more than 90% of the females had attended college, but more

were college graduates and fewer held advanced degrees. A much larger proportion of the females (60%) reported majors in mathematics, with no large grouping in any other field. All but 5% of the sample reported some kind of formal training in programming, the average amount being 160 hr. As might be expected, the proportion of females with machine-operations experience (20%) was much smaller than that for men (37%), and the average amount of operations experience was much less. Mean programming experience (5.3 yr.) of the female sample, however, was nearly identical to that of the male sample, despite the inclusion of a few 1-yr. programmers in the female group.

Employment characteristics of the female programmers were substantially different from those of the male programmers. Greater proportions of the females were in government and research and development organizations, and smaller proportions were in business and industrial firms. More of the females were on large programming staffs, and fewer were doing business programming. Fewer females were involved in requirements analysis, program system design and installation, and supervision. Median monthly salaries of the women ($757) averaged about $75 less than the men, despite the fact that experience of the two groups was about the same and about one-third of the women were responding 2 yr. later than the men. Examination of similar data for the portion of the female sample obtained in 1964 indicated that the above differences between male and female programmers were not due to differences in sampling or dates of collection. Apparently there are differences in the kinds of programming employment entered by, or available to, female programmers.

Procedure. For construction of interest keys the primary sample was divided randomly into two parts: a criterion group of 200 cases, and a validation group of 62 cases. The reference group represented an equal sampling from each of the 26 female SVIB occupational criterion groups.[1]

Items available for use in a new key for women included only those that:

[1] Reference-group response percentages were obtained from the Center for Interest Measurement Research, University of Minnesota.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF FEMALE PROGRAMMERS ON SVIB KEYS FOR MEN

Key	*M*	*SD*	Key	*M*	*SD*
1 Optometrist	37.53	11.51	31 Soc. Science Teacher	27.89	10.58
2 Physical Therapist	31.00	11.36	32 Sch. Superintendent	28.62	11.60
3 Dentist	32.71	10.59	33 Minister	26.23	10.99
4 Osteopath	28.80	09.65	34 Librarian	40.50	10.38
5 Veterinarian	18.59	09.71	35 Artist	31.24	10.31
6 Physician	37.85	11.64	36 Art Teacher	28.74	11.06
7 Psychiatrist	36.74	10.30	37 Musician Performer	44.38	09.24
8 Psychologist	37.10	10.46	38 Music Teacher	34.90	10.48
9 Biologist	38.84	11.19	39 C.P.A. Owner	37.61	09.10
10 Architect	36.81	11.12	40 Senior C.P.A.	40.32	08.84
11 Mathematician	36.00	10.73	41 Accountant	37.80	10.06
12 Physicist	29.51	13.31	42 Office Worker	35.64	09.98
13 Chemist	41.22	11.56	43 Credit Manager	32.12	11.00
14 Engineer	39.43	10.74	44 Chamber of Commerce Executive	31.10	10.37
15 Production Manager	34.27	07.46	45 Bus. Educ. Teacher	30.91	11.22
16 Army Officer	26.89	12.15	46 Purchasing Agent	26.11	09.36
17 Airplane Pilot	28.96	10.69	47 Banker	22.71	09.09
18 Carpenter	20.99	11.86	48 Pharmacist	29.95	08.45
19 Forest Service Man	22.11	12.32	49 Mortician	26.63	08.53
20 Farmer	29.60	09.49	50 Sales Manager	23.93	09.05
21 Indus. Arts Teacher	25.54	10.03	51 Real Estate Salesman	28.59	07.08
22 Math-Science Teacher	40.17	08.86	52 Life Ins. Salesman	22.30	09.20
23 Printer	30.49	08.95	53 Advertising Man	31.61	08.88
24 Policeman	24.95	07.75	54 Lawyer	33.62	08.48
25 YMCA Phys. Director	21.81	11.64	55 Author-Journalist	33.24	07.91
26 Personnel Director	31.64	11.78	56 Pres.-Mfg. Concern	20.02	09.20
27 Public Administrator	36.32	09.58	57 Specialization Level	44.33	09.66
28 Rehabilitation Couns.	31.30	10.19	58 Occupational Level	59.84	06.52
29 YMCA Secretary	20.85	11.93	59 Masculinity–Femininity	37.99	08.85
30 Social Worker	29.46	11.98	60 Computer Programmer	20.34	10.17

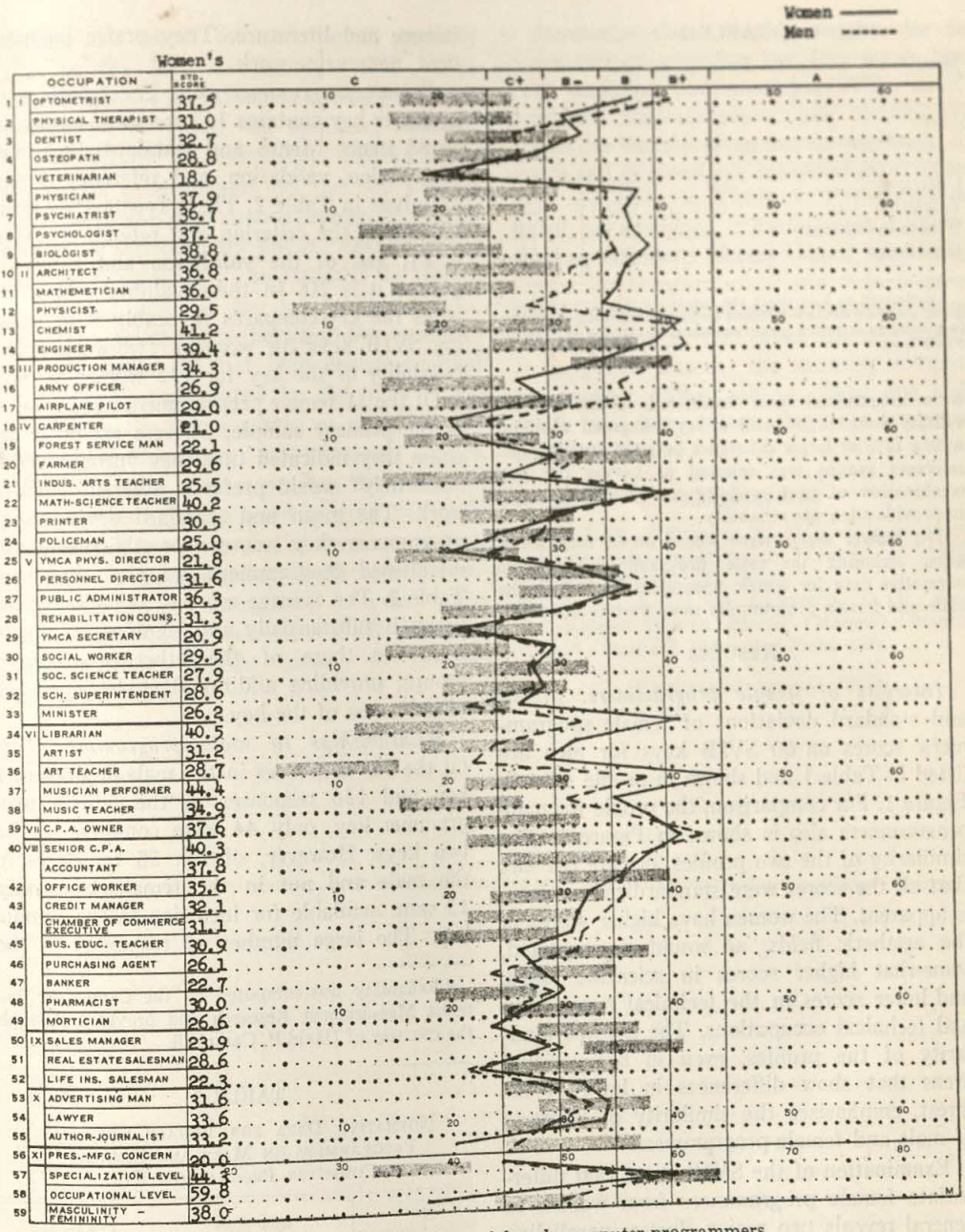


FIG. 1. SVIB profile for male and female computer programmers.

(*a*) had women-in-general response counts available (i.e., were in the 1946 edition of the SVIB for women); and

(*b*) appeared in the revised SVIBs for both men and women in identical or equivalent forms. Judgments of equivalence were rather conservative, and only a few items were not identical. These included 6 that involved direct changes in gender (e.g., actor-actress) and 18 that differed slightly in wording (e.g., member of supreme court—supreme court justice; president of society—president of society or club; author of "best seller"—author of best-selling novel).

There were 197 items that met these requirements and were therefore available for key construction.

Several interest keys of varying numbers of items

TABLE 2

MEANS AND STANDARD DEVIATIONS OF FOUR GROUPS ON THE FEMALE COMPUTER PROGRAMMER KEY

Group	*N*	*M*	*SD*
Criterion	200	20.26	11.52
Validation	62	18.21	9.38
Women in General		−11.52	13.73[a]
Dissatisfied	24	9.04	13.66

[a] Estimated as 1.24 times the combined criterion-validation group standard deviation, based on the average ratio of women-in-general to criterion group standard deviations for existing SVIB keys.

were constructed and evaluated. Differences in validity were slight, and a key composed of items with a 12% or larger difference between criterion and reference groups was selected as having the best combination of both validity and sufficient length to provide adequate reliability.

The answer sheets were scored on 60 male SVIB keys, including the male programmer key, and normative data for female programmers on both the male and female programmer keys were computed.

RESULTS

Interests of female programmers. Means and standard deviations of female programmers' scores on 60 SVIB keys for men are listed in Table 1 and shown in profile form in Figure 1. For comparison, the profile for male programmers also is shown in Figure 1. The similiarity of the two profiles, despite the fact that all the scores were standardized on men, is apparent. The women have higher scores in the aesthetic fields, as would be expected, somewhat higher scores in scientific fields, and lower scores in the technical supervision and technical occupations. The marked similarity of the profiles, even in occupational areas that show differences in level of interest, emphasizes the similarity of interests of male and female programmers.

Examination of the SVIB *items* that differentiate female programmers from women in general reveals two major clusters paralleling those for male programmers. These clusters indicate a strong interest in all forms of mathematics and a lack of interest in people —especially in activities involving responsibility for helping people. Female programmers are also more interested than other women in physical science and technical activities and less so in natural and social science and literature. They prefer independent, nonroutine work.

Female programmer key. The female programmer key contains 156 weighted responses to 88 items. Means and standard deviations of criterion, validation, and reference groups are given in Table 2. The percentage of overlapping of the criterion and reference groups is 21, and of the validation and reference groups it is 20. In these validity characteristics the key compares favorably with existing SVIB keys for women. Two-week retest reliability of the key for 112 women is .90.[2]

Of the 31 female programmers not included in the primary sample, 24 were excluded because they indicated that they "probably" or "definitely" would prefer a different field of work. The mean and standard deviation of programmer key scores for this group of dissatisfied programmers are also shown in Table 2. The average score of the dissatisfied group is substantially and significantly lower than are those of the other programmer groups, providing additional evidence of the effectiveness of the key.

Relationships to male programmer key. Of the 102 responses in the male programmer key and 156 responses in the female programmer key, only 44 were common to the two keys. However, of the 58 responses in the male and not in the female key, only 15 were available for inclusion in the female key. The large number of responses in the

[2] Reliability was computed by the Center for Interest Measurement Research and provided through the courtesy of David P. Campbell.

TABLE 3

NORMATIVE DATA FOR FEMALE COMPUTER PROGRAMMERS ON MALE AND FEMALE COMPUTER PROGRAMMER KEYS

Letter grade	Female standard score	Raw score		Percentage distribution on female key	
		Female key	Male key	Programmers	W-I-G[a]
A	45 to	14 to 81	16 to 56	70.2	3.4
B+	40 to 44	9 to 13	12 to 15	16.4	3.8
B	35 to 39	3 to 8	8 to 11	7.2	8.2
B−	30 to 34	− 2 to 2	4 to 7	3.4	10.1
C+	25 to 29	− 8 to −3	0 to 3	1.9	15.8
C	to 24	−77 to −9	−42 to −1	0.8	58.7

[a] Women-in-general estimated.

female and not in the male key is partially due to the less stringent criterion (12% rather than 15% difference between criterion and reference group) for including items in the female key. The correlation between scores of female programmers on the two keys is .60. The correlation among women in general would be substantially higher. Correcting for the larger variance to be expected for women in general increases the correlation to about .70. The relationship between the two keys is quite high, but not high enough so that one may be substituted for the other or the results on the two accepted as equivalent. If responses of women are obtained on the SVIB for men, their scores on the computer programmer key, expressed in terms of female norms, will be a meaningful measure of their interest in programming. Such scores, however, cannot be interpreted in exactly the same way as scores on other SVIB keys for men (without female norms) or on the SVIB for women.

Normative data. Scoring weights for the computer programming key for women have been provided to the major SVIB scoring services. Standard-score and letter-grade norms for female programmers on both the men's and the women's computer programmer keys are given in Table 3. The norms are based on combined criterion and validation groups. Standard scores based on female norms for the men's key are 4 to 5 points higher than standard scores based on men's norms, indicating that when the men's SVIB is used with women, the resulting computer programmer scores will be about one letter grade too low unless they are converted by Table 3.

REFERENCES

PERRY, D. K., & CANNON, W. M. Vocational interests of computer programmers. *Journal of Applied Psychology*, 1967, 51, 28–34.

STRONG, E. K. JR., & CAMPBELL, D. P. *Strong Vocational Interest Blanks (Manual)*. Stanford, Calif.: Stanford University Press, 1966.

(Received February 13, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 1, 36-41

SELECTION OF SALESMEN BY MEANS OF AN ASSESSMENT CENTER

DOUGLAS W. BRAY AND RICHARD J. CAMPBELL

American Telephone and Telegraph Company, New York City

Newly hired candidates for sales positions were evaluated by means of an assessment center consisting of paper-and-pencil tests, an interview, and individual and group simulations. Assessment staff judgments were compared with job performance some months later as evaluated by a special observational team. Assessment results were strongly related to this criterion. Supervisors' and trainers' ratings were not significantly related to the job-performance criterion nor to assessment results. The findings lend support to recent studies indicating the efficacy of the assessment-center method in personnel selection.

The assessment center (Taft, 1959) received its first industrial application when it was introduced into the Bell System in 1956 as a major research method of the Management Progress Study (Bray, 1964). Two years later, in 1958, an assessment center was opened in the Michigan Bell Telephone Company for the appraisal of candidates for promotion to management from vocational occupations. The results of this first operational assessment center were deemed so useful by line managers that the assessment-center method spread widely through the Bell System until over 50 assessment centers were established in the Bell System Companies processing approximately 8,000 men and women each year who are candidates for management positions.

Follow-up investigations in the Management Progress Study (Bray & Grant, 1966) demonstrated considerable predictive power for the assessment-center method both for young noncollege management men who had emerged from the ranks at an early age and for college recruits assessed very shortly after employment. In addition, a study of 500 men recently appointed to management (Campbell & Bray, 1967), some of whom had not been assessed and some of whom had received various ratings at the assessment center, further demonstrated the usefulness of the method.

The above assessment centers were designed to assess the ability to function successfully as a supervisory line manager. The present article describes the application of the assessment-center method to a quite different problem—the selection of prospective communications consultants. The job of the communications consultant is to work with business customers to determine their objectives, uncover inconsistencies in methods of reaching these objectives, and suggest improved communications, where appropriate, for improved operations.

A first step in devising an assessment center is a statement of the qualities relevant to success in the job for which assessment is planned. In order to arrive at this list of qualities, a psychologist experienced in the assessment-center approach due to his work on the Management Progress Study, David Berlew, spent several weeks interviewing sales personnel, observing sales training, and making numerous customer visits with experienced communications consultants. When a preliminary list of qualities had been isolated, assessment techniques designed to reveal these qualities were selected or developed. A group of experienced assessors, under the direction of the senior author, then applied these methods to a trial group of six successful salesmen.

This pretest resulted in some redefinition of the assessment qualities and some revision of the assessment-center techniques. The list of qualities was reduced to 20 and included such characteristics as resistance to premature judgment, oral presentation, oral defense, behavior flexibility, and persistence. The major assessment techniques were three simulations:

Leaderless Group Discussion—This tech-

nique involves six assessees. They are told to assume that they are attending a meeting with five of their peers to discuss candidates for promotion from among their fictitious subordinates. Each is given a description of a subordinate whom the assessee is to present as his candidate. Each assessee is given up to 5 min. to present his man, followed by a 1-hr. group discussion which must result in a rank ordering of the six candidates.

Oral Fact-Finding Exercise—This is an individual technique in which the assessee is told that he is an arbitrator in a labor-management dispute. He is given a short outline of the dispute and the opportunity to question a staff member to determine additional facts and issues. He then presents his decision and its rationale orally to the staff member and is questioned intensively about it.

Consulting Case—This is an individual technique in which the assessee is given written material—textual, graphic, and tabular—concerning the operation and problems of a business concern. His task is to study this material and prepare recommendations for the company. He presents his recommendations orally to two staff members who role play the two controlling partners of the enterprise and discusses his recommendations with them. He then prepares a final written recommendation.

In addition to these simulations, the assessees completed a biographical data blank, underwent a lengthy interview, and took four tests, all cognitive in nature. These tests were the School and College Ability Test, a Critical Thinking Test, an Abstract Reasoning Test, and a Test of Knowledge of Contemporary Affairs.

A trial of the assessment-center methods outlined above was conducted by setting up a Sales Assessment Center in Cleveland, Ohio to which some 14 participating Bell Telephone Companies sent candidates with little or no Bell System sales experience. This assessment center was staffed by a seven-man team, including a fourth-level director and six third-level sales managers from various telephone company sales departments. This staff received 3 weeks' training by the authors and others before beginning the actual assessment of the trial *S*s.

Assessees were processed by the assessment center in groups of six, and it required 2 full days for each group of six to undergo the various assessment techniques. Only one group of six was processed per week, however—the remaining days of each week being used by the assessment staff in preparing reports of the behavior of each candidate in each exercise and for holding staff sessions to hear these reports for each candidate, to rate him on the 20 sales-assessment qualities, and to make a final recommendation as to his suitability as a communications consultant.

In all, some 142 men from the 14 participating telephone companies were assessed, but various factors reduced the sample of men for the present report to 78. This sizable reduction was due to the fact that not all of the men sent to the assessment center met the standards which had been set up for the study. Some had had previous Bell System sales experience, others were on a special training program for college recruits, and some were not assigned to communications consultant work after assessment. In addition, a few resigned from the company before on-the-job criteria were collected.

Great care was taken that the results of the man's performance at the assessment center did not affect his assignment or appraisal on the job. All of the men in the study were Bell System employees at the time of their assessment and there was no feedback to their trainers, their supervisors, or to the men themselves on their performance at the assessment center. It was anticipated that if the assessment center proved a valuable screening method, it would be used as a pre-employment screen, but since the individual results of the trial assessments were not to be fed back to the supervisors immediately concerned with the man's training or assignments, all *S*s were employed. In fact, the participating companies were urged to keep as many as possible on the payroll so that criterion data could be collected after they had had sales training and were on the job for several months.

Before proceeding to a discussion of the criterion against which the assessment judgments were compared, it may be of interest to examine the final judgments of the candi-

TABLE 1

ASSESSMENT-STAFF JUDGMENTS OF ACCEPTABILITY FOR SALES EMPLOYMENT

Assessment judgment	No. men	Percentage of group
More than acceptable	9	12%
Acceptable	32	41%
Less than acceptable	16	20%
Unacceptable	21	27%
Total	78	100%

dates made by the assessment staff. After the reports on each candidate were read at the weekly staff meeting and the 20 sales qualities rated on a 5-point scale, the assessment staff made a final judgment as to each candidate's acceptability as a potential communications consultant. The 78 assessees were put into four groups—one group was called "more than acceptable," another "acceptable," a third group was considered "less than acceptable," and the final and lowest group was considered "unacceptable." Table 1 shows the distribution of assessment-staff judgments. It will be seen that only 9 of the 78 men performed outstandingly at the assessment center while 21 men did very poorly. The top two groups—the acceptable groups—totaled 41 men, or 53% of the sample, while 47% were considered unacceptable in varying degrees.

This finding itself is of interest. It indicates that even though the 78 assessees had met all employment standards and had been screened by their local companies as qualified for the job of communications consultant, experienced sales managers judged that from what they could see at the assessment center, only about half possessed good potential for their proposed job. This result is in keeping with other assessment findings in the Bell System, both of the college recruits of the late 1950s and the male and female candidates for management in the various telephone companies. Assessment staffs characteristically reject one-half or more, usually more, of the candidates.

The distribution of assessment-staff judgments was not significantly different for the 64 men who were eliminated from the study. Forty-seven percent of that group "passed" the center as compared to 53% of those retained in the sample.

The primary criterion against which assessment-center judgments were compared was first-hand observation of actual behavior in sales contacts 6 mo., on the average, after assessment. By that time each of the candidates had had sales training and had been working in the field for several months. This criterion had been decided upon before the assessment study was even begun. It was considered more pertinent than "sales results" which are influenced by many factors other than the behavior of the salesman. Company sales executives had, in any case, indicated that because of differences in market characteristics in the many areas across the United States to which these candidates were assigned, no statistic or group of statistics would accurately reflect individual performance.

The performance observations were carried out by a field-review team working out of American Telephone and Telegraph headquarters in New York. This group's regular work involved such field reviews, and it was an entirely experienced team that went out on these special criterion reviews. A reviewer accompanied each *S* on as many visits as necessary to determine whether he did or did not meet established standards in conducting his sales activities. These standards included preparation, usage prospecting, recommendations, closing, and implementation. As a result of the observations, each man's performance was classified as meeting standards or failing to meet standards. The reviewers were, of course, entirely unaware of any judgments

TABLE 2

ASSESSMENT JUDGMENTS AND FIELD-PERFORMANCE RATINGS

Assessment judgment	No. men	No. meeting review standards	Percentage meeting review standards
More than acceptable	9	9	100%
Acceptable	32	19	60%
Less than acceptable	16	7	44%
Unacceptable	21	2	10%
Total	78	37	47%

Note.—$\chi^2 = 24.19$; $p < .001$.

or observations made of the candidate at the assessment center.

Of the 78 men, 37 men or 47% met review standards. Table 2 shows the relationship between the assessment-center judgments and whether or not the salesmen met the standards. It will be seen that all of the 9 men deemed more than acceptable at the assessment center met review standards, while only 2 of the 21 men deemed completely unacceptable at the assessment center passed the field review.

A supplementary investigation[1] looked into the ratings of the *Ss* given by their sales supervisors and by the trainers who had had each man in sales training. A preliminary letter was sent to each supervisor indicating that he would be asked to rate each salesman reporting to him. He was asked to prepare himself to do this and, when ready, to telephone either of a two-man team in the sales staff in A.T.&T. in New York. The New York staff member verified that he and the supervisor each had the same list of salesmen, and the supervisor was then asked to rank all of the men reporting to him in terms of their excellence as a communications consultant. After the ranking was recorded, the supervisor was then asked to rate each man on a graphic rating scale, the upper end of which was labeled "highest—I wish I had many more like him" and the lower end labeled "lowest—He's in the wrong job." The supervisor was then asked whether he knew if any of the men had been assessed and, if he did, he was asked whether he knew how the man had come out at the assessment center. In most cases, the supervisors did not know the men had been assessed and in no case did the supervisor have any accurate information about the outcome of the man's assessment. The same procedure was carried on with trainers, who rated and ranked all men in each training group in which an assessee had been trained.

[1] This approach was devised by Ira Cisin of George Washington University.

TABLE 3

ASSESSMENT JUDGMENTS AND SUPERVISORS' RATINGS

Assessment judgment	No. men	Supervisors' rating	
		Higher	Lower
Acceptable	41	57%	43%
Less than acceptable	37	47%	53%

Note.—Differences not statistically significant.

TABLE 4

SUPERVISORS' RATINGS AND FIELD-PERFORMANCE RATINGS

Supervisors' rating	No. men	Met review standards	Did not meet review standards
Higher	41	47%	53%
Lower	37	44%	56%

Note.—Differences not statistically significant.

In order to maximize differentiation for purposes of comparison, the average rating given by the supervisors and by the trainers was determined. The study group was then divided at about the midpoint of each set of ratings. In the following tables, those men who scored below the midpoint are described as having received higher ratings and those who scored below the midpoint are described as having received lower ratings. It should be noted, however, that the ratings were biased in an upward direction, so some of the men counted as receiving lower ratings were not, in an absolute sense, rated below average.

Table 3 shows the relationship between the assessment-center judgments and the supervisors' ratings. It can be seen that there was not a strong relationship—57% of those who received acceptable assessment-center judgments received higher ratings by the supervisors, as compared to 47% of those who had received unacceptable assessment-center ratings. This difference is not statistically significant.

These results immediately raised the question of to what extent the supervisor's ratings were related to the field-performance ratings. Table 4 shows the results. It will be seen that there was very little or no relationship between supervisory ratings and ratings of performance on the sales job by the field-review team. Such a result has, of course, consider-

TABLE 5

ASSESSMENT JUDGMENTS AND TRAINERS' RATINGS

Assessment judgment	No. men	Trainers' rating	
		Higher	Lower
Acceptable	41	51%	49%
Less than acceptable	37	55%	45%

Note.—Differences not statistically significant.

TABLE 6

TRAINERS' RATING AND FIELD-PERFORMANCE RATING

Trainers' rating	No. men	Met review standards	Did not meet review standards
Higher	41	56%	44%
Lower	37	40%	60%

Note.—Differences not statistically significant.

able import for questions of appraisal. Of considerable methodological importance, however, is the observation that the assessment center might have been considered not sufficiently accurate for use if supervisory judgment had been relied upon as the sole criterion.

Table 5 shows the relationship between the assessment-center judgments and the trainers' ratings. Here the relationship is very slight and negative, if indeed there is a relationship. There is somewhat more of a suggestion that trainers might be able to predict field performance, as Table 6 shows. Fifty-six percent of the men who received higher training ratings met review standards as compared to 40% of those who had received lower ratings. This difference is, however, not statistically significant. In addition, 85% of the *S*s had been given above-average training ratings—not a very fine screen.

In spite of this lack of strong evidence for the usefulness of trainers' ratings, there remained the possibility that such ratings could make a further selection of those who were deemed acceptable (or better) by the assessment center. Table 7 shows the relationship between trainers' ratings and field performance for the 41 men who were judged acceptable (or better) by the assessment staff. There is only a very slight, and insignificant, relationship between trainers' ratings and performance, and almost two-thirds of those who get lower ratings show good performance. It does not appear that the trainers' ratings, as they were made in this study, would be of practical use as a selection device. There remains, of course, the possibility that trainers' observational and rating methods could be improved.

It may be of interest to compare the degree of prediction obtained from the assessment-center staff judgment and that from the four paper-and-pencil tests. Table 8 shows the data in correlation form. It will be seen that the highest correlation of any of the tests with the criterion is .28. The multiple correlation using all four tests is .33 (not corrected for shrinkage). These, of course, are considerably below the .51 correlation of the assessment judgment with the field-performance ratings.

The biographical blank used at the assessment center was administered to give the staff members some advance information be-

TABLE 7

TRAINERS' RATING AND FIELD-PERFORMANCE RATING (MEN JUDGED ACCEPTABLE OR BETTER AT ASSESSMENT CENTER ONLY)

Trainers' rating	No. men	Met review standards	Did not meet review standards
Higher	21	71%	29%
Lower	20	65%	35%

Note.—Differences not statistically significant.

TABLE 8

CORRELATIONS OF ASSESSMENT PREDICTION AND PAPER-AND-PENCIL TESTS WITH FIELD-PERFORMANCE RATINGS

Assessment judgment	.51
SCAT	.25
Critical Thinking Test	.26
Contemporary Affairs Test	.28
Abstract Reasoning Test	.02
Multiple *R*—Four Tests	.33

fore conducting interviews with the candidates. The blank was not scored. Because of the small size of the sample, it is unwarranted at this time to analyze these biographical items or to attempt to combine by multiple correlational techniques all the various pieces of data gathered during the 2 days of assessment. Such analyses can be done later when more cases accumulate and are now almost completed for a management assessment center (Bray & Grant, 1966; Grant, Katkovsky, & Bray, 1967).

DISCUSSION

The assessment-center method of evaluating potential for various occupations has had a spotty history (Taft, 1959). Conclusions about its effectiveness have been difficult to draw because the number of applications of the method has been small, techniques and procedures have varied markedly from one application to the next, and criterion problems are formidable. Studies of Bell System assessment centers for the prediction of success in management positions have, however, uniformly shown that the method has substantial validity. This has been true whether the assessment staff consisted primarily of professional psychologists or of specially trained management personnel. The present study demonstrates that an assessment center staffed by sales managers instructed in assessment techniques can afford a valuable aid in the selection of prospective salesmen.

REFERENCES

BRAY, D. W. The management progress study. *American Psychologist*, 1964, **19**, 419–420.

BRAY, D. W., & GRANT, D. L. The assessment center in the measurement of potential for business management. *Psychological Monographs*, 1966, **80**(17, Whole No. 625).

CAMPBELL, R. J., & BRAY, D. W. Assessment centers: An aid in management selection. *Personnel Administration*, 1967, **30**(2), 6–13.

GRANT, D. L., KATKOVSKY, W., & BRAY, D. W. Contributions of projective techniques to assessment of management potential. *Journal of Applied Psychology*, 1967, **51**, 226–232.

TAFT, R. Multiple methods of personality assessment. *Psychological Bulletin*, 1959, **56**, 333–352.

(Received February 20, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 1, 42-48

A BIOGRAPHICAL INVENTORY FOR IDENTIFYING CREATIVITY IN ADOLESCENT BOYS[1]

CHARLES E. SCHAEFER AND ANNE ANASTASI

Fordham University

A biographical inventory was developed and cross-validated in a sample of 400 high school boys in the New York metropolitan area, subdivided into creative and control groups in artistic and scientific fields. The principal criterion was teachers' nominations supported by specific creative products. Validity coefficients in the cross-validation sample were .64 and .35 for artistic and scientific keys (each $p < .001$). Final keys, containing items differentiating in initial and cross-validation samples (compound $p < .05$), were used to describe similarities and differences between creative students in the 2 fields.

Among the many approaches followed in current investigations of creativity, a major distinction is that between the use of test scores and of evaluated achievement as criteria of creativity. The former approach is illustrated by the research of Guilford (1959, 1967), Getzels and Jackson (1962), and Torrance (1962); the latter by that of Roe (1951a, 1951b, 1953) and MacKinnon (1962). With tests as a criterion, not only is coverage likely to be limited to certain aspects of creative behavior, but motivational variables are also largely excluded.

As a predictor and as an instrument for exploring the correlates of creativity, the biographical inventory offers particular promise, having demonstrated its validity in predicting many other complex criteria in industrial, military, and educational contexts (Henry, 1966). When cross-validated against criteria of scientific creativity, biographical inventories have proved successful with several types of research workers. Relevant investigations include those of Smith, Albright, Glennon, and Owens (1961) with petroleum research scientists, Buel (1965) with research personnel in a pharmaceutical company, McDermid (1965) with technical engineering personnel, and Chambers (1964) with eminent psychologists and chemists. In a series of investigations of National Aeronautics and Space Administration (NASA) scientists, Taylor, Ellison, and Tucker (1966) obtained validity coefficients in the .40s and .50s when biographical inventories were cross-validated against several criteria of creative achievement.

With only a few exceptions, investigations of creativity at the high school or college level have employed tests as criteria, predictors, or both. Dauw (1966) reports successful differentiation between highly creative and less creative adolescents by means of a biographical inventory, but his subjects (*Ss*) were selected solely on the basis of creativity test scores. Parloff and Datta (1965) compared contrasted groups of participants in the Westinghouse Science Talent Search, selected on the basis of judges' ratings of their research projects. However, these groups were compared chiefly in personality test scores, the only background items reported being father's occupation, socioeconomic level, and intactness of family. Taylor, Cooley, and Nielson (1963) applied a modified version of the biographical inventory developed on NASA scientists to high school students participating in a National Science Foundation (NSF)-supported summer science program. The biographical inventory proved to be the best overall predictor of the creative research performance of these students, its validity reaching .47 in one of the groups. In a follow-up of National Merit Scholarship Finalists, Nichols and Holland (1963) correlated biographical data with self-reported evidence of creative achievement during the

[1] The present study is part of a larger project supported initially by Grant No. MH 10233-01 from the National Institute of Mental Health and subsequently by Subcontract No. 2 of the Center for Urban Education's Contract OEC-1-6-062868-2039 with the United States Office of Education.

first year of college. The relationships found were rather low, a fact that may be attributable in part to inadequate criterion data and narrowly restricted ability range of the sample.

The present study [2] was concerned with the development and cross-validation of biographical inventory keys against achievement criteria of creativity among high school boys. An examination of these keys provides a description of significant antecedents and correlates of creativity in this population.

METHOD

Subjects

The *Ss* employed in the principal data analyses were 400 male students from six high schools in the New York metropolitan area.[3] The group included 188 seniors, 144 juniors, and 68 sophomores. Socioeconomic level and parental education of the entire sample were superior for the geographical area. The schools were chosen because they offer courses or programs providing an opportunity for creative achievement.

Creativity Criteria

Four criterion groups, of 100 students each, were formed as follows: Creative-Artistic (CrA), Control-Artistic (CoA), Creative-Scientific (CrS), and Control-Scientific (CoS). The *creative Ss* had to meet two criteria: (*a*) teacher nomination on the basis of one or more creative products to be described on a teachers' nomination form; (*b*) score above a minimum cutoff on Guilford Alternate Uses and Consequences tests. The "artistic" category covered graphic art and literary expression; the "scientific" included sciences and mathematics.

The control *Ss* were nominated by the same teachers as having produced no evidence of creative achievement, and they scored below a maximum cutoff on the two Guilford screening tests. It should be noted that the Guilford tests were employed only as a check on irrelevant factors that might have influenced the nomination of creative or control *Ss*. The scores were used only to exclude cases—never to admit them. Moreover, the two cutoff scores were sufficiently extreme as to exclude only individuals whose test performance was highly discordant with their reported achievement. In terms of available published norms, the mean scores of the creative students on the two Guilford tests were approximately equal to those of college students, while the mean scores of the control groups approximated those of ninth-grade students.

The control *Ss* were *matched* with the creative *Ss* in school attended, class, and grade-point average (GPA). Since the CrS students significantly excelled the CrA students in GPA, two separate control groups were assembled. Although no attempt was made to equate groups in socioeconomic level, no significant differences were found either between creative and control or between artistic and scientific groups in mean Shevky-Bell index—an index based on the characteristics of the census tracts in which the individual resides.

Biographical Inventory

A comprehensive 165-question biographical inventory was prepared for this study. The questions were formulated principally on the basis of previous research findings of other investigators and hypotheses regarding the correlates of creativity. The 165 questions were grouped into five sections labeled: physical characteristics, family history, educational history, leisure-time activities, and miscellaneous. While most of the questions dealt with objective facts regarding present or past activities and experiences, some called for expressions of preference and others concerned anticipated plans and goals.

Although some multiple-choice and checklist items were included, many questions were open-ended. Even with the objective items, there was provision for adding other unlisted responses. While scoring and data analysis are more difficult under these conditions, these types of items yield a richer return of information and are especially appropriate in an exploratory study. All responses were coded prior to tabulation. For each question, there were several possible responses, and for some questions this number was quite large. In addition, several questions yielded responses that could be classified from different viewpoints to test different hypotheses. Consequently, the 165 questions yielded a total of 3,930 "scorable items" or individual response alternatives used in the item analyses.

Procedure

The biographical inventory (together with three tests used in another part of the project) was administered to groups of 30–180 students during a 2-hr. testing session held outside of school hours. The *Ss* were paid for participating in this testing session. Identification numbers were employed to provide anonymity, and students were assured of the confidentiality of the data.

For purposes of data analysis, each of the four groups was subdivided into two groups of 50, used for development of scoring keys and cross-validation, respectively. Each pair of groups was equated in number of students from each school, class distribution, GPA, and mean score on the screening

[2] More detailed information on procedure and results can be found in Schaefer (1967).

[3] The authors acknowledge the invaluable cooperation of J. Wayne Wrightstone, assistant superintendent, Board of Education of the City of New York, and of the administrative and teaching personnel of the participating schools: Forest Hills, Jamaica, Midwood, and Stuyvesant (public high schools in New York City), Bronxville (a suburban public high school), and Regis (a private high school conducted by the Jesuit order in New York City).

TABLE 1

BIOGRAPHICAL INVENTORY SCORES OF CRITERION GROUPS IN CROSS-VALIDATION SAMPLES

Score	Creative-Artistic key[a]		Creative-Scientific key[b]	
	Creative-Artistic	Control-Artistic	Creative-Scientific	Control-Scientific
Over 110	6	0	0	0
101–110	6	1	0	0
91–100	11	2	0	0
81–90	10	4	1	0
71–80	7	8	7	2
61–70	8	2	15	3
51–60	2	7	9	11
41–50	0	9	5	9
31–40	0	9	6	13
21–30	0	3	4	10
11–20	0	3	2	2
1–10	0	2	1	0
N	50	50	50	50
M	88.76	52.38	53.50	41.62
SD	19.58	24.32	17.69	13.91
Range	52–140	3–110	4–84	17–72
NCR	8.23*		3.74*	
$r_{p_{bis}}$	.64*		.35*	

[a] Possible range −96 to 335.
[b] Possible range −82 to 239.
* $p < .001$.

tests. For each of the 3,930 scorable items, classified as present or absent, a phi coefficient was computed against the dichotomous criterion of creative versus control. These coefficients were computed separately in artistic and scientific criterion groups. All items with phi coefficients at the $p < .20$ or better significance level were considered for inclusion in the *initial* CrA or CrS scoring keys. Some of these items were excluded because they duplicated other items, were checked by three or fewer *S*s, or were inconsistent with other responses and hence were likely to have yielded isolated chance correlations.

In the initial scoring keys, a weight of 1 was assigned to items discriminating between the $p < .20$ and $p < .05$ levels, and a weight of 2 to items discriminating at the $p < .05$ level or better. Items with higher frequencies in the creative group received positive weights; those with higher frequencies in the control group received negative weights. The initial CrA and CrS scoring keys were used in scoring the biographical inventories of the corresponding creative and control *S*s in the cross-validation samples. The scorers were unaware of the criterion status of the *S*s. The scores thus obtained were correlated with the dichotomous criterion to provide an estimate of the validity of the scoring keys. In order to utilize all the data in the selection of items for *final keys,* item analyses were carried out independently in initial and cross-validation samples and those items were selected that differentiated between criterion and control groups with a compound probability of .05 or better (Baker, 1952; Katzell, 1951).

RESULTS

The cross-validation data for the initial keys are summarized in Table 1.[4] In both artistic and scientific groups, the mean scores of creative *S*s excelled those of controls at the .001 level of significance, with normal curve ratios of 8.23 and 3.74, respectively. The point-biserial criterion correlations, whose significance is indicated by the above-normal curve ratios, were .64 and .35, respectively.

Examination of the frequency distributions in Table 1 indicates that for both keys a cutoff score of 60 yields the maximum differentiation between creative and control groups. A simple index of differentiation for this purpose is the difference between cumulative percentage frequencies in creative and control groups. With a cutoff above 60, 96% of the CrA *S*s in the cross-validation sample will be correctly classified, while 34% of the controls will be incorrectly classified as creative. With the CrS key, 46% of the creative *S*s will be correctly classified, while 10% of the controls will be incorrectly classified as creative. As proposed by Levy (1967), the "substantive significance" of a group difference may also be evaluated in terms of the percentage of misclassification, defined as the percentage of individuals deviating from the mean of their own group more than halfway toward the mean of the other group. This percentage, which can be readily found from the point-biserial correlation, is 21 for the CrA key and 36 for the CrS key, as compared to a chance expectancy of 50.

Once the validity of the initial keys (derived from one-half the available cases) was established, the final keys were constructed with items whose compound probability was based on data obtained from the total sample. The CrA key thus developed contains 140 scorable items, while the CrS key contains 65. Examination of the items in these final

[4] In order to avoid negative scores in computation, 70 was added to each score on the Artistic key and 40 to each score on the Scientific key. This adjustment, however, did not exclude negative scores from the total possible range, as shown in the Table 1 footnotes.

keys provides the most conclusive available description of the characteristics of the creative students surveyed in this investigation; this description will be summarized in the next section.

DISCUSSION

Identifying Creativity in Artistic versus Scientific Fields

In all comparisons—including mean differences, criterion correlations, extent of overlap of score distributions, and number of significantly differentiating items in initial and final keys—the CrA students were more sharply differentiated from their controls than were the CrS students. It should be recalled in this connection that, in both groups, creative and control *Ss* were equated in GPA and that CrS students significantly excelled CrA students in this regard. From many sources of evidence it appears that the CrS high school student resembles the academically superior student more closely than does the CrA student. Both the nature of scientific creativity itself and current educational practices in science courses suggest that high academic achievement—together with its characteristic antecedents and correlates—plays a predominant role in scientific creativity. On the other hand, the traditional activities of artists and writers, as well as the content of courses in these areas, differ quite conspicuously from typical academic activities. While creative students in both artistic and scientific fields can be significantly differentiated from academically equated controls by means of a biographical inventory, this study strongly suggests that the former can be more readily differentiated than the latter.

Correlates of Creativity across Both Fields

An examination of items common to the final CrA and CrS keys provides a picture of the correlates of creativity in high school boys, regardless of field. In terms of specific scorable items, only 14 differentiated in the same direction at the .05 level or better in both final keys. This number is somewhat of an artifact, however, since there were many other differently worded but basically equivalent items in the two keys. Based upon both identical and equivalent items, the following description focuses upon the best established characteristics which were corroborated by several items. It might be added that even these common characteristics tended to be revealed in more items in the CrA than in the CrS group.

The *familial background* of the creative students was not only academically superior, but the parents tended also to provide role models of interest and creative expression in the student's field. College education was more frequent among parents of creative students, especially among mothers. Magazines regularly available at home were more likely to be of the political commentary-foreign affairs or the cultural-intellectual types. Creative students were less likely than controls to come from homes in which neither student nor parents play a musical instrument. With regard to specific field of interest, both fathers and mothers in the CrA group read more books than those in the control group, and reading was more often listed as the father's favorite leisure-time activity. The fathers had more often won honors or awards in literary or art fields. The mothers had more often specialized in art or literary subjects in college, had more creative hobbies, more frequently visited art museums or galleries, and were more often listed as the family member who had taught the student most about art. CrA students were also less likely than controls to come from families having no hobbies. The CrS students more often than the controls reported that science magazines were read regularly in the home and that their fathers had scientific hobbies. It is also relevant to note that in the CrS group one or more siblings were likely to have received national recognition for academic achievements.

In a number of ways, the creative student himself manifests a *strong intellectual and "cultural" orientation.* In both artistic and scientific groups, creatives more often than controls reported that they were considering an Ivy League college. In both groups, a significantly greater number of creatives reported that they read over 50 books a year. Other relevant items, occurring in either artistic or scientific groups, pertain to types of news-

papers and sections of newspapers read regularly, membership in academic and honor societies, and interest in plays, music, and art museums. Both creative groups, on the other hand, reported significantly less interest in sports than did their respective control groups, as evidenced by attendance at spectator sports, reading about sports, or even daydreaming about them.

A third set of items can be characterized as indications of *pervasive and continuing enthusiasm* for one's chosen field and singleness of purpose in its pursuit. Early manifestations of creativity are evident in several items, such as childhood creation of new games, mechanical or electronic objects, poems, stories, or art products. Persistence of interest in the same field is indicated in elementary school awards and in favorite and best subjects in elementary school. Creative students more often than controls received awards in their special field outside of school and they were more likely to own such field-related articles as microscopes, chemistry sets, and artists' supplies. They also reported having become so absorbed in a project in their field that they skipped a meal or stayed up very late. They more often chose hobbies in their special fields, which were usually directly related to their vocational goals. Extracurricular activities, membership in out-of-school organizations, magazines read, and summer jobs all illustrate the pervasiveness of their interest. The creative adolescent evidently finds his chosen work so satisfying and refreshing that he is willing to devote his leisure time to it. He chooses an area that is personally and deeply valuable to him and is hence able to persevere at it.

The creative student's perseverance and goal orientation are not to be construed as narrowing influences, however, since they are likely to be accompanied by a characteristic *breadth of interest*. Thus creative students in both fields were more likely than their controls to have pursued a variety of hobbies (frequently since childhood), to have collections of different sorts, to play muscial instruments, to attend plays, and to participate in extracurricular activities. In such extracurricular activities, moreover, they were more likely to have held offices and to have been elected to the highest offices, such as president or editor. Another example is provided by the CrA students who, when asked to check their favorite high school subjects, checked more subjects than did the controls.

Finally, certain items suggest that creative students exhibit a stronger drive toward *novelty and diversity* in their activities than do controls. To some extent, the previously discussed breadth of interest may be regarded as a manifestation of this drive. But other illustrations can be cited. For example, creative students reported more frequent preference than controls for unusual art forms, such as Pop, Op, and primitive art; they had more unusual collections, such as road signs and spider webs; and they more often owned different kinds of records, such as classical and folk-song. Similarly, the tendency of creative students in both fields to prefer friends both older and younger than themselves may indicate a desire for diversity in interpersonal relations.

Differences between Creativity Correlates in Artistic and Scientific Groups

A direct comparison of the biographical inventory responses of CrA and CrS students would be contaminated by differences in school and class distribution and GPA. In the present experimental design, each creative group was equated with its own control group in these variables. Hence the question now to be considered is whether the characteristics that differentiate the CrA students from their own controls differ in any systematic way from those that differentiate the CrS students from their controls. This question can be answered by comparing the two final biographical inventory keys.

The most conspicuous difference between the CrA and CrS keys occurs in a set of items indicative of the CrA student's greater exposure to *environmental diversity*, through either personal or parental experiences. The parents of the CrA students were more likely to have been born in a different city, state, or country than the one in which the family was living. In the CrS group, the reverse relation was found, the parents being *less* likely than those in the control group to have been born in a foreign country and more likely to have

been born in the home city. CrA students were also more likely than controls to have moved from one state to another during their lifetime and to have visited more different parts of the country and more distant parts than the controls; they had also more often traveled out of the country—to Mexico, the Caribbean, and South America.

Several items in the CrS key give evidence of some *conventionality* in both parents and students, evidence that is certainly absent from the CrA keys. The CrS students were more likely than their controls to have fathers who majored in business subjects in college, who play bridge as a hobby, and from whom the students had learned about sports. The CrS students themselves participate in individual-type high school sports such as tennis more than do the controls, although no significant group difference was found in team sports. In contrast, CrA students spend less time participating in sports of any kind than do their controls. These and other items also suggest a stronger sex-role identification among the CrS than among the CrA students, the former tending to enjoy closer relationship with the father, the latter with the mother.

With regard to *childhood experiences*, the CrS students reported less evidence of social participation in early childhood than did the controls. They more often stated, for example, that they had only one close friend and were never a leader of their childhood group activities under the age of 12. Although the picture of early childhood that emerges from the CrA key does not contrast with that of the CrS key, it nevertheless highlights a somewhat different pattern of experiences and traits. Thus the CrA child was more likely than the controls to daydream and to have imaginary companions; his earliest childhood memory was more likely to date back to the age of 4 or earlier; and he had more unusual experiences, such as eidetic-type imagery. His relationships to his parents appear to have centered more often on shared interests in artistic or literary pursuits and less often on a warm personal relationship than was true of the control *S*s. For example, the CrA student more often described his mother as enthusiastic and less often as understanding or sincere. He attended nursery school more often and did not describe family discipline as "always fair" as frequently as did the controls. If he had a personal problem, he was more likely to seek help from "several persons" than from his mother.

A few noteworthy differences emerged with regard to *school experiences*. Preference for artistic and literary activities appeared in the CrA students' reactions to their elementary-school curriculum, in questions pertaining to favorite subjects and subjects in which the students did best or poorest. In contrast, no items relating to the elementary-school curriculum differentiated CrS students from controls, these differences first appearing at the high school level. This finding may indicate that elementary-school instruction offers more creative opportunities in artistic and literary fields than in science. Both groups of creative students gave some differentiating responses to questions about outstanding teachers, but the focus of their replies differed. The CrA students reported having had two or more such teachers more often than did the controls. The CrS students differed from the controls in their more frequent listing of "wide knowledge" as the principal characteristic of such teachers and also in naming more than one trait to characterize them. It is also of some interest to note that CrS students tended to give more "uncertain" answers in the biographical inventory as a whole, a finding that may be related to the habits of *caution* fostered by scientific training.

In conclusion, creative students in artistic and scientific fields do share important common features in experiential background and personal traits, but they also exhibit notable differences in both respects. Consequently, any investigation of creativity that fails to consider area of creative achievement may produce a blurred or distorted picture of the creative person. Although the distinction between scientific and artistic creativity is undoubtedly a major one, further breakdown of artistic creativity into literary production and graphic art might be fruitfully explored in further research.

REFERENCES

BAKER, P. C. Combining tests of significance in cross-validation. *Educational and Psychological Measurement*, 1952, 12, 300–306.

BUEL, W. D. Biographical data and the identification of creative research personnel. *Journal of Applied Psychology*, 1965, 49, 318–321.

CHAMBERS, J. A. Relating personality and biographical factors to scientific creativity. *Psychological Monographs*, 1964, 78(7, Whole No. 584).

DAUW, D. C. Life experiences of original thinkers and good elaborators. *Exceptional Children*, 1966, 32, 433–440.

GETZELS, J. W., & JACKSON, P. W. *Creativity and intelligence*. New York: Wiley, 1962.

GUILFORD, J. P. Three faces of intellect. *American Psychologist*, 1959, 14, 469–479.

GUILFORD, J. P. *The nature of human intelligence*. New York: McGraw-Hill, 1967.

HENRY, E. R. Conference on the use of biographical data in psychology. *American Psychologist*, 1966, 21, 247–249.

KATZELL, R. A. Cross-validation of item analyses. *Educational and Psychological Measurement*, 1951, 11, 16–22.

LEVY, P. Substantive significance of significant differences between two groups. *Psychological Bulletin*, 1967, 67, 37–40.

MACKINNON, D. W. The nature and nurture of creative talent. *American Psychologist*, 1962, 17, 484–495.

MCDERMID, C. D. Some correlates of creativity in engineering personnel. *Journal of Applied Psychology*, 1965, 49, 14–19.

NICHOLS, R. C., & HOLLAND, J. L. Prediction of the first year college performance of high aptitude students. *Psychological Monographs*, 1963, 77(7, Whole No. 570).

PARLOFF, M. B., & DATTA, L.-E. Personality characteristics of the potentially creative scientist. *Science and Psychoanalysis*, 1965, 8, 91–106.

ROE, A. A psychological study of eminent biologists. *Psychological Monographs*, 1951, 65(14, Whole No. 331). (a)

ROE, A. A psychological study of physical scientists. *Genetic Psychology Monographs*, 1951, 43, 121–235. (b)

ROE, A. A psychological study of eminent psychologists and anthropologists, and a comparison with biologists and physical scientists. *Psychological Monographs*, 1953, 67(2, Whole No. 287).

SCHAEFER, C. E. Biographical inventory correlates of scientific and artistic creativity in adolescents. Unpublished doctoral dissertation, Fordham University, 1967.

SMITH, W. J., ALBRIGHT, L. E., GLENNON, J. R., & OWENS, W. A. The prediction of research competence and creativity from personal history. *Journal of Applied Psychology*, 1961, 45, 59–62.

TAYLOR, C. W., COOLEY, G. N., & NIELSEN, E. C. Identifying high school students with characteristics needed in research work. National Science Foundation Grant 17543, University of Utah, 1963. (Mimeo)

TAYLOR, C. W., ELLISON, R. L., & TUCKER, M. F. *Biographical information and the prediction of multiple criteria of success in science*. Greensboro, N. C.: Richardson Foundation, 1966.

TORRANCE, E. P. *Guiding creative talent*. Englewood Cliffs, N. J.: Prentice-Hall, 1962.

(Received February 23, 1967)

Journal of Applied Psychology
1965, Vol. 12, No. 1, 49-54

AUTOMOBILE ACCIDENTS AND DRIVER REACTION PATTERN [1]

PAUL BABARIK [2]

Accident Research Center, American Institutes for Research, Washington, D. C.

Taxicab drivers who reported an abnormal number of accidents in which they were struck from behind, when tested on a classical laboratory reaction-time apparatus, were found to have a reaction pattern made up of slow initiation time and compensatingly fast movement time. Drivers with this perceptual-motor pattern probably stopped their vehicles more abruptly in a way that cannot be duplicated by a following driver, or in the headway of the following vehicle. However, drivers with such an atypical perception reaction had fewer accidents in their headways and therefore had lower overall accident rates. Suggestions for further research and the possible implications of these findings were discussed.

Motor-vehicle accidents are not a homogeneous class of events with a common set of causes but heterogeneous events with multiple causes. In spite of this, much accident research uses such criterion data as total number of accidents during the previous 5 yr., or overall accident and violation rates. This practice may have contributed to the limited progress in accident research. Focusing on one normal aspect of driving such as stopping, and one kind of accident, the struck-from-behind accident, should result in greater research success. Although driving is a perceptual-motor skill, few relationships between perceptual-motor variables and accident criterion data have been found (e.g., Goldstein, 1962; Lauer, 1959). The abandonment of perceptual-motor variables does not necessarily follow from these findings. Why not consider perceptual-motor variable patterns or combinations and relate these to the behavioral patterns of drivers that might lead to accidents? Drake (1940) considered perceptual-motor tests in combination and found that individuals who performed faster on motor tests than perceptual tests had more industrial accidents than individuals whose perceptual speed was faster than their motor speed. He hypothesized that this perceptual-motor discrepancy should identify high-accident drivers.

Although the existence of either a general perceptual speed factor or of a general motor speed factor has been disproved (e.g., Fleishman, 1953, 1954), the possibility that a refined pattern of component perceptual-motor factors should be associated with specific driving acts, if not overall driving success, has to be considered as an alternative. Of course, such hypothesized predictor variables must be related to more refined driving measures than overall accident rate. Accordingly, the primary aim of this study was to determine the relationship between one pattern of perceptual-motor behavior and one type of accident. The perceptual-motor behavior considered was the ratio of simple reaction time (RT) to jump reaction time (JRT) and the type of accident was the rear-end collision or the struck-from-behind accident.

The struck-from-behind accident was chosen as the criterion variable not only because of the common-sense likelihood of a link with reaction pattern, but also because this kind of accident is numerous enough to yield sufficient cases for statistical analysis; for example, in one recent state study this class of accident accounted for 25% of all highway accidents, and what is of even greater significance, 30% of all injuries (State of New York, 1964). In freeway driving, where

[1] This study was carried out under Grant No. AC 00222-01, awarded by the Division of Accident Prevention, United States Public Health Service to the Washington Office of the American Institutes for Research.

[2] Now with Computer Applications, Inc., Silver Spring, Maryland.

there are no intersections and all traffic is moving in the same direction, more than 50% of the accidents fall into this category.

In the present study, response time on reaction-time tests had been fractionated into initiation time and movement time. Initiation time is the time required to begin the movement in response to the stimulus. The time to complete the perception reaction as represented in the jump reaction time (JRT) test is made up of initiation time plus movement time.[3]

In the present study, it was hypothesized that drivers whose initiation time is an abnormally long portion of their total perception reaction (i.e., drivers who have a high ratio of RT to JRT) would have a higher percentage of accidents in which they were struck from behind than would drivers whose perceptual-motor RT shows a typical ratio. Why should there be any difference in the stopping characteristics of two drivers, both of whom take .60 sec. for their total perception reaction, but who differ in the length of the initiation time and movement components? Why should there be a difference between Driver A who typically takes .15 sec. to initate the response and .45 sec. for movement, thus has a reaction-pattern score of $\frac{.15}{.60}$ or .25, and Driver B who takes .30 sec. for movement, thus has a reaction-pattern score of $\frac{.30}{.60}$ of .50? A vehicle's stopping is partly a function of the force that is applied by the driver, and force equals Mass × Acceleration (F = MA). Therefore, it was hypothesized that the driver who makes up for long initiation time with fast movement, that is, has short movement time, has greater limb acceleration and therefore applies greater force and stops his vehicle in fewer feet after contact with his brake pedal is made and his brake lights come on. Actually with such compensating limb speed, not only is the acceleration increased, but the mass is increased as additional *g* forces enter into braking to make the stop even more abrupt and unrepeatable by a following driver. As total reaction time is transformed into reaction distance, drivers with this reaction pattern typically stop in less road distance or within the expected headway of the following driver. By systematically leaving less distance between themselves and following drivers, such drivers increase their probability of being struck from behind, especially when such hazards as tailgating and slippery roads exist. In effect, long initiation time followed by compensating short movement time probably desynchronizes the traffic flow or constitutes a desynchronizing reaction pattern.

The struck-from-behind accident is a prima-facie extreme instance of such a reaction pattern. While the principal hypothesis was that drivers with a desynchronizing reaction pattern will have a higher incidence of struck-from-behind accidents than drivers with normal reaction patterns, if the atypical drivers stop short, the corollary hypothesis that they should collide with vehicles in their headway *less* frequently would also follow and was also tested. Although struck-from-behind accidents constitute a large proportion of accidents, the majority of accidents still occur in one's headway. Therefore, it would follow that drivers with a desynchronizing reaction pattern would have a lower overall accident rate.

The question might well be raised as to why the probability of a struck-from-behind accident is not primarily a function of a simple perceptual-motor factor movement time (as measured by JRT − RT), rather than as a function of a less parsimonious and more tricky ratio, the desynchronizing reaction pattern (as measured by RT/JRT). First, it is not more parsimonious to use movement time, as movement time is also a derived score, obtained by subtracting RT from JRT. Second, to use movement time as a predictor of the accident criterion would be to ignore previous findings of no relationships between movement time and accidents (Goldstein, 1962). Third, as noted above, Drake (1940) found a relationship between perceptual-motor ratio and accident proneness

[3] The author is indebted to A. Zavala for this information on the RT and JRT tests. These components have been identified in factorial studies (e.g., Fleishman, 1954) as the reaction-time factor (RT) and the speed of limb movement factor (SLM). Reference to these components as initiation time and movement time is consistent with both factorial findings and logical analysis.

but none between individual perceptual-motor tests and accidents. Finally, a relationship between the desynchronizing reaction pattern and automobile accidents, especially struck-from-behind accidents, follows from a rational consideration of automobile braking. The braking perception reaction approaches the laboratory JRT only in a forced or emergency stop. The driver whose initiation time is usually longer will have to utilize his shortest movement time more often than the driver with faster initiation time. While it is granted that movement time determines the stopping distance, other things being equal, it seems reasonable to assume that short movement time will increase the probabilities of struck-from-behind accidents only when it follows a long initiation time that necessitates a "crash" stop. Accordingly, it was also hypothesized ad hoc that movement time alone would not correlate with accident incidence. Since measures of RT and JRT were to be obtained in computing the desynchronizing-reaction-pattern and the movement-time factors, possible relationships between these two variables and struck-from-behind accidents could also be examined. Following Goldstein (1962) it was also hypothesized ad hoc that no relationship between the RT and JRT tests and accident incidence exists.

METHOD

The data upon which this study was based were obtained from 127 Washington, D. C. taxicab drivers. The study began with 160 *Ss*. However, only the 127 experienced cab drivers—those with 2 or more yr. full-time hacking or equivalent (part-time hacking was considered as one-half time)—were retained. The *Ss* averaged 48 yr. of age. Each *S* was asked for the number of accidents he had had in the past 5 yr. He was then asked to describe each accident. Drivers were first classified according to their overall accident rates into three groups: (*a*) 0 accidents or *accident free*, (*b*) *1 accident*, and (*c*) 2–9 or *multiple accidents*. Then those drivers who reported one or more accidents were classified according to kind of accident—struck-from-behind and other-than-struck-from-behind (Other), or accidents in their *tailway* and accidents in their *headway*.

The drivers were administered two kinds of RT tests. One was simple reaction time (RT) where *S*'s finger rested on a microswitch which was depressed in response to a light signal. The second test was jump reaction time (JRT) where *S* had to move his hand 12 in. from a rest position to the button in making the response. In both tests, there was a set of 20 trials separated by a fore-period ranging from .5 to 1.5 sec.—varied in random order. The ratio of RT to JRT was computed and the drivers were classified as having either a normal or an atypical reaction pattern. A ratio that was .5 standard deviations above the overall mean was considered to be great enough to constitute an atypical or desynchronizing reaction pattern (DRP). Contingency tables were prepared in which the accident rates of the Normal and DRP groups were compared. Measures of initiation time—(JRT minus RT)—and of jump reaction time (JRT) were also computed and correlated ad hoc with accident incidence.

RESULTS

To study the relationship of the reaction pattern to kinds of accidents, the 23 *Ss* reporting no accidents were removed since this group contributed no variance regarding their kind of accident; or, to be fatalistic, they had not manifested their kinds of accidents yet.

From Table 1 it appears that drivers with the atypical reaction pattern, if they had an accident, were more apt to be struck from behind than were drivers who have a normal reaction pattern. The drivers with a desynchronizing reaction pattern were heavily represented in the multiple struck-from-behind accident category—36% of the 28 drivers found to have this pattern had been struck from behind more than once, or were chronic struck-from-behind accident "victims" com-

TABLE 1

COMPARISON OF TAXICAB DRIVERS WITH A DESYNCHRONIZING REACTION PATTERN (DRP) AND NORMAL REACTION PATTERNS WHO HAD ONE OR MORE ACCIDENTS WITH RESPECT TO STRUCK-FROM-BEHIND (SFB) ACCIDENTS

	Reaction pattern		Total no.
	Normal	DRP	
No. SFB accidents			
0 (only Other) accidents	37	9	46
1 accident	31	9	40
2–8 multiple	8	10	18
Total	76	28	104

Note.—$\chi^2 = 9.19$; $p < .02$.

TABLE 2

COMPARISON OF TAXICAB DRIVERS WITH A DESYNCHRONIZING REACTION PATTERN (DRP) AND NORMAL REACTION PATTERNS WITH RESPECT TO MULTIPLE STRUCK-FROM-BEHIND (SFB) AND MULTIPLE OTHER ACCIDENTS

	Reaction pattern		Total no.
	Normal	DRP	
Multiple (2–8) SFB accidents	8	10	18
Multiple (2–8) Other accidents	32	6	38
Total multiple-accident drivers	40	16	56

Note.—$\chi^2 = 9.86$; $p < .001$.

pared with 10% of the 76 drivers with a normal reaction pattern.

While a single accident may be a chance event, multiple accidents of the same type present unusually strong support for the operation of a systematic variable and causal link. So, drivers with two or more accidents were compared further. The double-edged effect of the atypical reaction pattern is shown in Table 2. Only 6 out of 16 (or 38%) of the multiple-accident DRP group had multiple Other accidents, whereas 32 (80%) of the 40 drivers with a normal reaction pattern had their accidents "up front." The corollary hypothesis that drivers with the desynchronizing reaction pattern have fewer Other accidents was thus supported.

Having considered the relationships of the reaction pattern to accidents in one's tailway and also in one's headway, the safe driver—one who had neither class of accidents or was accident-free—was examined next. In Table 3 the accident-free drivers were compared with drivers who had either headway or tailway accidents. Thirty percent of the 40 drivers with a desynchronizing reaction pattern were found to be accident free compared to only 13% out of the 87 drivers with normal reaction patterns. When the number of accidents was examined, the Normal group reported 2.01 accidents in the preceding 5 yr., whereas the DRP group reported 1.65 accidents over the same period. On two-way roads with cross traffic the majority of accidents occur in one's headway. The Other accident rate reflects this fact. However, in one-way driving without intersections the struck-from-behind accident is often the prevailing kind of accident. When most driving occurs under such circumstances, then the driver with the desynchronizing reaction pattern may well have the highest overall accident rate and no longer be the safer driver.

Only the desynchronizing-reaction-pattern score accounted for accident variance when chi-square contingency tables were used. When product-moment correlation was utilized to compare the four possible variables, no relationship was found between the single

TABLE 3

COMPARISON OF TAXICAB DRIVERS WITH A DESYNCHRONIZING REACTION PATTERN (DRP) AND NORMAL REACTION PATTERNS WITH RESPECT TO OVERALL ACCIDENTS

No. accidents	Reaction pattern		Total no.
	Normal	DRP	
0	11	12	23
1	32	9	41
2–9	44	19	63
Total	87	40	127

Note.—$\chi^2 = 6.35$; $p < .05$.

TABLE 4

CORRELATION OF REACTION-TIME VARIABLES AND STRUCK-FROM-BEHIND (SFB) ACCIDENTS

	r
Initiation time (IT) and no. SFB accidents	.13
Jump reaction time (JRT) and no. SFB accidents	.05
Movement time (JRT − IT) and no. SFB accidents	.23*
Desynchronizing reaction pattern $\left(\frac{RT}{JRT}\right)$ and no. SFB accidents	.29**

* $p < .05$.
** $p < .01$.

perceptual-motor scores of RT and JRT and accident incidence. Contrary to hypothesis, movement time was significantly correlated with struck-from-behind accidents. However, the finding that the correlation of movement time was lower than the correlation of desynchronizing reaction pattern with struck-from-behind accidents could be taken as support for the general thesis that while the speed of movement time determines the stop, the stop becomes more hazardous when movement time is combined with a slow initiation time as represented by the desynchronizing reaction pattern.

DISCUSSION

The finding of an empirical relationship, if confirmed by further studies, has practical significance. However, considerable research is called for before it would be clear how this finding would pay off. The crux of such research would be to test the assumption underlying the desynchronizing reaction pattern, that is, that the DRP driver's perception reaction in stopping results in systematically less space between him and the following car. The arrangement used in research on car following (Gazis, Herman, & Rothery, 1961) could measure the hypothesized differential space relationship resulting from the desynchronizing reaction pattern and other variables considered in this present study. If the hypothesized hazard is confirmed and specified quantitatively, further research to determine whether the desynchronizing reaction pattern is relatively fixed or immutable would be called for. Implications along the following lines might follow if the desynchronizing reaction pattern were found to be relatively stable:

(*a*) *Selection and driver testing*. Taxicab drivers, chauffeurs, patrolmen, and commercial drivers could have a test of their reaction pattern included in their selection tests and the results put on record.

(*b*) *Training*. Follow-up counseling and instruction procedures could be tried and, if found effective, incorporated into driving training in general. Specific counseling that one is prone to a specific kind of accident, in this case to being struck from behind, might prove more effective than general warnings to be careful.

(*c*) *Legal*. This measure could constitute evidence for assigning a greater portion of the costs and culpability to a leading driver. This might not only be more equitable than the present tendency to charge the following driver, but also provide an incentive to the antecedent driver to compensate for his desynchronizing reaction pattern. Such a driver might also have a warning fixed to the rear end of his car after he had a struck-from-behind accident to caution following drivers.

(*d*) *Human engineering*. If the perceptual-motor regularity studied here is established as a variable in accidents, changes in the braking system might be called for. These might even take such forms as dampening the application of braking force for some drivers, more effective signaling of stopping to the following driver, and perhaps adjusting the braking system to the driver as one adjusts the seats.

If this pattern is mutable, then changing the desynchronizing reaction pattern would be called for. The most obvious way would be to reduce the initiation-time factor and thus bring the perception-reaction pattern into line with typical drivers. This would assume that the pattern has a large attitudinal rather than psychophysiological basis. One possible and feasible way of altering the reaction pattern is to slow down the movement-time component so that it is no longer compensatingly fast.

A slowing of movement to decrease a hazard and probability of accident seems incredible. Would not a reduction of accidents in one's tailway be at the expense of more accidents in one's headway? A post hoc examination of the 40 DRP drivers brought to light 6 men who had both very long RT and long JRT; that is, their movement time, while short in relation to their initiation time, was longer than the sample mean or not fast in absolute terms. It was found that this group had a mean struck-from-behind accident rate of .33 (total sample mean = .74), with only one driver reporting any struck-from-behind accidents; he reported two. This group reported a total of five accidents, or a

mean rate of .83 (total sample mean = 1.83). Three of these six drivers were accident free. While no conclusions are warranted on the basis of this small group alone, the hypothesis, that a slowness in initiating a response without compensating limb speed is associated with the below-average accident rate, calls for further investigation. It may be that these drivers learned to live with their slow reaction pattern. If subsequent studies were to confirm that such a reaction pattern does not lead to more headway accidents, and results in a lower rate of tailway accidents, in a roundabout way an aspect of safe driving would be identified.

REFERENCES

DRAKE, C. A. Accident proneness: A hypothesis. *Character and Personality*, 1940, **8**, 335–341.

FLEISHMAN, E. A. Testing for psychomotor abilities by means of apparatus tests. *Psychological Bulletin*, 1953, **50**, 241–262.

FLEISHMAN, E. A. Dimensional analysis of psychomotor abilities. *Journal of Experimental Psychology*, 1954, **48**, 437–454.

GAZIS, D. C., HERMAN, R., & ROTHERY, R. W. Nonlinear follow-the-leader models of traffic flow. *Operations Research*, 1961, **9**, 545–567.

GOLDSTEIN, L. G. Human variables in traffic accidents. *Highway Research Board Bibliography 31*. National Academy of Science, National Research Council, Washington, D. C., 1962.

LAUER, A. R. A comparison of group paper and pencil tests with certain psychophysical tests for measuring driver aptitude for Army personnel. *Journal of Applied Psychology*, 1959, **39**, 318–321.

STATE OF NEW YORK, DEPARTMENT OF MOTOR VEHICLES. Motor vehicles accidents caused by following too closely. January to October 1963. *Statistical Bulletin No. 7*, 1964.

(Received February 27, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 1, 55–61

FACULTY SATISFACTION AND THE DEAN'S INFLUENCE: AN ORGANIZATIONAL STUDY OF TWELVE LIBERAL ARTS COLLEGES[1]

JERALD G. BACHMAN

Institute for Social Research, University of Michigan

Questionnaire data were obtained from 685 faculty members in 12 liberal arts colleges. Measures included amounts of administrative influence, bases of influence for the dean and faculty, and levels of faculty satisfaction. Faculty satisfaction was greatest in those colleges having the highest total influence across all levels of the academic hierarchy. Faculty also indicated greatest satisfaction with deans who have relatively high influence over college affairs, and whose influence is based upon expertise and respect rather than upon legitimate authority and coercion. The results are discussed in the light of similar findings in business organizations.

The research reported in this paper has two basic purposes. The first is to provide some tentative, but hopefully useful, guidelines for college administrators. A second and broader purpose is to explore the parallels between administrative processes in higher education and those in other kinds of organizations, especially industrial ones. Studies of this sort should guide our efforts to apply the results of organizational research to the academic setting. And perhaps more important, they should tell us something about the possibilities for developing general theories of administration, applicable to business, governmental, political, voluntary, and academic organizations. The present paper, which focuses on the role of the academic dean, is the first of several reports based on this research in liberal arts colleges.

In many respects, the academic dean is the faculty's most important voice in the administration of the liberal arts college. He represents the faculty viewpoint to the president and trustees, and he also represents higher administration to the faculty. His role is thus similar to many middle-management positions in business organizations. Given this set of dual loyalties and commitments, what are some of the qualities of an effective dean? If for the moment we are guided by research findings in nonacademic organizations, the following hypotheses may be considered:

1. The effective dean is influential. A study in industry by Pelz (1951, 1952) indicated that a supervisor's effectiveness in dealing with his subordinates depends in part upon his influence with his own superiors; the influential supervisor can "deliver the goods" for his men. A recent study of account executives engaged in the sale of intangibles (Bachman, Smith, & Slesinger, 1966) produced similar results; satisfaction and performance tended to be higher when the office manager was perceived as highly influential over "how the office is run." To the extent that these findings are applicable to an academic situation, they imply that the "strong" dean will be effective.

It does not follow that if the dean is highly influential then someone else in the college administration will be proportionately less influential. Recent studies suggest that the total amount of influence (or control) in an organization is best treated as a variable, not a fixed value; indeed, a relatively high level of control across all organizational echelons is associated with high performance and

[1] The research reported here was supported by grants from the Carnegie Corporation of New York to the Survey Research Center and the Center for Study of Higher Education, The University of Michigan.

Data for the present analysis are taken from a larger survey carried out by A. L. Fisch and the author. Special thanks are due to the many faculty members who devoted their time and effort to provide the information for this study.

Thanks are also due to the following colleagues who read an earlier draft and offered suggestions: Robert Cope, Glenn Jones, Arnold Tannenbaum, George Wieland, and Alvin Zander.

satisfaction (Tannenbaum, 1962). Tannenbaum (1961) and Likert (1961) have argued that this pattern of "high total control" is successful because it involves members at all levels of the organization, leading to more effective decisions and also to higher motivation. This suggests a supplementary hypothesis: the effective college administration is one with relatively high influence at *all* levels.

2. The effective dean's influence is based upon a relatively high degree of competence and personal attractiveness, and a relatively low dependence upon his "legitimate authority" and coercion. Five different "bases of power," originally distinguished by French and Raven (1960), were examined by Bachman et al. (1966). Satisfaction and to some extent performance were positively related to "expert" power (based on competence and experience) and "referent" power (based on personal admiration); negative correlations occurred with "legitimate" power (based on hierarchical "rights") and "reward" and "coercive" power (based on the potential use of positive and negative sanctions).

The above hypotheses are stated in very broad terms. For example, they deal with the dean's "effectiveness" as if it were a unitary dimension. In fact, of course, there may be many criteria for evaluating performance in the role of dean. The present research is necessarily limited to studying the dean's effectiveness in terms of his impact on the faculty; specifically, the criteria will be several measures of faculty satisfaction, including their view of how well the dean fulfills his responsibilities.

METHOD

Research Sites

The respondents selected for study consisted of all full-time faculty members at each of 12 liberal arts colleges belonging to a regional association of colleges. This choice provided a number of institutions which are somewhat homogeneous with respect to size, geographic location, and educational reputation. Since our interest is concentrated upon the effects and correlates of certain administrative factors, it seemed desirable to select colleges which would be fairly similar in other characteristics. A second and more practical reason for the choice was the willingness of member colleges to participate in the study.[2]

At the time of the survey (April–May 1964), faculty size at the 12 colleges ranged from 67 to 173, with a median of 92 and a mean of 101. A total of 1,210 questionnaires were sent to eligible faculty members; after several months, during which time two follow-up letters and a duplicate questionnaire were sent to each nonrespondent, a total of 685 were returned. While a higher response rate would have been preferred, of course, the final response (approximately 60% in each of the 12 colleges) was consistent with the author's expectations for a long (19-page) mail questionnaire administered near the end of the academic year.[3]

Measures

All data used in the study consist of faculty responses to questionnaire items. Only a small proportion of the total questionnaire material has been included in the present analysis.

Influence. The following question was used to assess influence over college matters: "In general, how much influence does each of the following groups or persons actually have in determining the policies and actions of your colleges?" Response categories ranged on a 5-point scale from (1) "Little or none," to (5) "A great deal." The respondents were asked to rate the amount of influence of each of the following groups or persons: the board of trustees, the president, the academic dean, the department chairmen (all of them), the faculty as a group, and the students as a group. The present study uses two measures based on this item: a measure of "total influence" (as indicated by the mean of all six influence ratings), and the influence of the academic deans (based on the rating of the dean alone).

Bases of influence. The following questionnaire item was used to assess five bases of the dean's influence over the respondent (referent, expert, reward, coercive, and legitimate influence, in that order):

> Listed below are five reasons generally given by people when they are asked *why* they do the things their superiors suggest or want them to do.

---

[2] The problem of gaining access to a research site and obtaining the cooperation of prospective respondents is crucial in research of this sort. Special thanks are due to Algo D. Henderson and Eldon L. Johnson, of the Center for Study of Higher Education, for their help in assuring the full cooperation of each college in the association.

[3] It must be stated clearly that the purpose of the research based upon this sample of college faculty is not to describe "the average faculty member." Rather, the primary intent is to discover *relationships* among variables; such relationships are likely to appear even if the set of respondents departs somewhat from an accurate probability sample drawn from the universe of faculty members.

Please read all five carefully. Then number them according to their importance to you as reasons for doing the things your academic dean suggests or wants you to do. Give rank "1" to the most important factor, "2" to the next, etc.

A. I respect him personally, and want to act in a way that merits his respect and admiration.

B. I respect his competence and judgment about things with which he is more experienced than I.

C. He can give special help and benefits to those who cooperate with him.

D. He can apply pressure or penalize those who do not cooperate.

E. He has a legitimate right, considering his position, to expect that his suggestions will be carried out.

An additional item was included on an exploratory basis to measure faculty perceptions of their influence *over the dean*. The respondents were given a list of alternatives similar to those listed above, but in this case they were asked to ". . . number them according to what you think is their importance to your academic dean as reasons for doing the things *you* suggest or request of him."

Because of the ranking procedure involved, the five bases of influence are not independent; in a sense, any single basis of influence can be given prominence only at the expense of the other bases. On the other hand, the ranking procedure has the advantage of forcing the respondent to discriminate among all the bases, rather than giving prominence to only one or two. Moreover, it avoids a common contaminating tendency to rate the *amount* of influence rather than the nature or bases of influence.

Faculty satisfaction. Three measures of satisfaction were used; two deal with overall "job satisfaction," and the third specifies satisfaction with the dean.

In order to obtain some indication of commitment or attachment to present college positions, each respondent was asked to indicate (*a*) his first choice of another college or university and (*b*) the minimum salary condition that would induce him to leave his present position to accept a similar one at the other college or university. Response alternatives covered a 7-point scale ranging from (0) "I would accept even if it involved about a 20% salary reduction," to (6) "I would not accept such a position, no matter how great a salary increase might be involved."

A second measure of satisfaction simply presented the respondent with the following statement, and asked him to indicate his agreement or disagreement on a 5-point scale: "While no job can be expected to be perfect, there are really very few things that I would change about mine if I had the power to do so."

The measure of satisfaction with the dean also presented a statement, and asked the respondent to indicate agreement or disagreement on a 5-point scale: "All things considered, I am personally quite satisfied with the way the academic dean of my college fulfills his responsibilities."

TABLE 1

COLLEGE ADMINISTRATIVE CHARACTERISTICS AND FACULTY CRITERIA: MEANS AND STANDARD DEVIATIONS

	M rating ($N = 685$)	*SD* among individuals ($N = 685$)	*SD* among college *M*s ($N = 12$)
Influence			
By dean	3.53	1.00	0.51
Total influence	3.49	0.47	0.14
Bases of dean's influence			
Legitimate	2.40	1.14	0.20
Expert	1.94	1.15	0.38
Referent	2.48	1.09	0.19
Reward	3.67	0.96	0.16
Coercive	4.40	1.07	0.26
Bases of faculty influence			
Legitimate	1.72	0.91	0.14
Expert	1.93	0.92	0.11
Referent	2.85	0.91	0.17
Reward	3.66	0.86	0.16
Coercive	4.78	0.63	0.13
Faculty satisfaction			
Minimum conditions for leaving	3.53	1.66	0.41
With job as a whole	3.29	1.11	0.19
With dean of college	3.56	1.13	0.45

Clearly, these several measures do not relate to a simple unitary dimension of satisfaction. It is quite possible, for example, to find a faculty member who would rate his present position favorably, and yet be eager to go to a large and prestigious university—even at a slight reduction in salary. And that same respondent could be either satisfied or dissatisfied with the way the dean fulfills his responsibilities. Nevertheless, moderate positive correlations appeared among the three measures. The product-moment correlation between the two measures of overall satisfaction is .35; satisfaction with the dean correlates .20 with attachment to present position, and .34 with the second measure of overall satisfaction (in each case $p < .01$, two-tailed test).

RESULTS

The major findings of the study are summarized in Tables 1, 2, and 3. Table 1 presents mean responses and standard deviations for all variables. Tables 2 and 3 present correlations between college administrative characteristics and faculty criterion ratings at two levels of analysis (described below). Before presenting the substantive findings, it will be helpful to outline the procedures and rationale for data analysis.

Statistical Analysis and Rationale

The analytic approach used in this study closely parallels that used by Bachman et al. (1966). This approach assumes that the best available measure of an "organizational characteristic" is often some composite of percep-

TABLE 2

COLLEGE ADMINISTRATIVE CHARACTERISTICS CORRELATED WITH MEAN FACULTY CRITERION RATINGS

College administrative characteristic	College mean faculty criterion ratings		
	Minimum conditions for leaving	Satisfaction with job as a whole	Satisfaction with dean of college
Influence			
By dean	.16	−.07	.69*
Total influence	.42	.19	.74**
Bases of dean's influence			
Legitimate	−.24	−.27	−.52
Expert	.38	.56	.75**
Referent	.46	.57	.67*
Reward	−.31	−.61*	−.80**
Coercive	−.44	−.57	−.70*
Bases of faculty influence			
Legitimate	−.16	.42	.47
Expert	.44	.61*	.62*
Referent	.23	.11	.39
Reward	−.40	−.75**	−.75**
Coercive	−.05	−.24	−.71**

Note.—N = 12 colleges.
* $p < .05$, two-tailed, product-moment correlation.
** $p < .01$, two-tailed, product-moment correlation.

tions by a number of respondents in the organization. Thus in the present study the measure of a college administrative characteristic always consists of the *mean* of all faculty perceptions. Each characteristic is correlated with faculty satisfaction measures at two distinct levels of analysis: college mean criterion scores and individual criterion scores.

College mean criterion scores. The first level of analysis is the college as a whole; Table 2 presents correlations computed between administrative characteristics and mean faculty satisfaction scores ($N = 12$ colleges). An individual faculty member's satisfactions with his position and dean are presumably influenced by (*a*) certain college-level effects—administrative characteristics which are common to most of his campus colleagues, and also (*b*) a host of idiosyncratic factors which differ from person to person. The use of mean satisfaction scores largely cancels the effects of personal factors (*b*), and thus highlights the impact of college-level effects (*a*).

Individual criterion scores. The second level of analysis is the individual faculty member; Table 3 presents correlations computed between administrative characteristics (still based on *mean* perceptions) and individual faculty satisfaction ratings ($N = 685$ faculty members). This form of analysis deals with the question: Given the many causes of any individual's satisfactions, what is the *relative* importance of certain college administrative characteristics? It is clear at a glance that the correlations in Table 3 are, on the whole, much smaller than those in Table 2. In Table 2 college-level effects have a "monopoly" in their impact upon satisfaction, since individual differences are held more-or-less constant through the use of college mean satisfaction data. In Table 3 this impact of college-wide factors is "shared" with the many causes that may be different for each individual.

Isolation of college-level effects. An additional question which can be treated at the

TABLE 3

COLLEGE ADMINISTRATIVE CHARACTERISTICS CORRELATED WITH INDIVIDUAL FACULTY CRITERION RATINGS

College administrative characteristic	Individual faculty criterion ratings		
	Minimum conditions for leaving	Satisfaction with job as a whole	Satisfaction with dean of college
Influence			
By dean	.05	.00	.26**
	−.02	−.08	.06
Total influence	.11*	.03	.27**
	.07	−.03	.20**
Bases of dean's influence			
Legitimate	−.06	−.07	−.21**
	−.05	−.06	−.18**
Expert	.09*	.10*	.29**
	.05	.05	.18**
Referent	.11*	.11*	.26**
	.09*	.08	.22**
Reward	−.07	−.11*	−.34**
	−.07	−.09*	−.31**
Coercive	−.11*	−.11*	−.27**
	−.08	−.06	−.20**
Bases of faculty influence			
Legitimate	−.04	.09*	.21**
	−.04	.09*	.20**
Expert	.09*	.09*	.24**
	.08	.08	.22**
Referent	.05	.02	.14**
	.04	.00	.10*
Reward	−.08	−.13**	−.31**
	−.07	−.10*	−.28**
Coercive	.00	−.01	−.27**
	.01	−.03	−.22**

Note.—N = 685 faculty members. For each cell in the table two correlations are given: The upper value is the zero-order correlation, r_{Ab}; the lower (boldfaced) value is the partial correlation, $r_{Ab.a}$; where,

A = college-wide mean of ratings of administrative characteristics
b = individual faculty criterion ratings (satisfaction, etc.)
a = individual faculty ratings of administrative characteristics.

As a matter of convenience, all significance levels were calculated on the basis of 500 *df*. Actual response varied somewhat from item to item, but no combination of items (*ab*) provided less than 500 *df*.
* $p < .05$, two-tailed, product-moment correlation.
** $p < .01$, two-tailed, product-moment correlation.

individual level of analysis has to do with the *way* in which administrative characteristics have an influence on individual satisfactions: Is it simply the *perception* of these characteristics which affects satisfaction? In order to deal with this issue, each relationship between an administrative characteristic and individual faculty satisfaction is presented in two forms in Table 3: the zero-order correlation and a partial correlation. The partial correlation rules out the effects of each individual's own perception of the college administrative characteristic, thus removing that portion of the relationship which might be attributed to "phenomenological" or "halo" effects. The portion that remains thus represents a "phenomenology-free" measure of the college-level effect. For a discussion of the background and rationale for this form of analysis, see Tannenbaum and Bachman (1964); its implications for the present study are treated later in this paper.

Influence

The mean level of influence attributed to the dean is midway between "moderate" and "considerable." As Table 1 indicates, the standard deviation among college mean ratings is relatively large; thus it appears that the 12 deans vary considerably in their influence. The mean total influence rating is also midway between "moderate" and "considerable"; however, the variation from college to college is much smaller (perhaps because influence in this case is averaged across six levels from board of trustees to students).

Colleges in which the total amount of influence is high, and those in which the dean's influence is high, show greater mean satisfaction with the dean. These administrative characteristics are also related to individual faculty members' satisfaction with the dean; however, the correlations with dean's influence disappear almost completely when the effects of individual perceptions are removed through partial correlations (Table 3). The more general measures of faculty satisfaction (minimum conditions for leaving, and satisfaction with the job as a whole) show little relationship to our influence measures at either level of analysis.[4]

Bases of Influence

When asked to rank reasons for doing what the dean suggests or wishes, faculty members gave the highest ranking to the measure of expert influence ("I respect his competence and judgment about things with which he is more experienced than I"); next most important were legitimate and referent influence; lowest rated were reward and coercive influence, that is, the use of sanctions (Table 1). Correlations with the criterion variables were positive for expert and referent influence, negative for legitimate, reward, and coercive influence.[5]

Turning next to faculty perceptions of why the dean follows faculty suggestions or requests, it is clear that the results are largely similar to those for the bases of the dean's influence. The important exception is legitimate influence. Faculty members indicated that the dean's most important reason for acting as they wish is that "The dean feels that his job properly includes an obligation to consider and act upon my suggestions as a faculty member." Moreover, this legitimate influence of the faculty over the dean is positively related to satisfaction with the dean. The ordering of the four other bases of dean's influence, and the actual mean rankings are quite similar (see Table 1). The correlations with satisfaction measures are

[4] It is of interest to note that the respondents were also questioned about the *ideal* amount of influence their dean should have. The level they recommended was consistently higher than their rating of the dean's actual influence. The increase occurred in 11 of the 12 colleges; in the twelfth case, the estimates of actual and ideal influence were practically identical, and very high.

[5] Some caution must be exercised in interpreting correlations with the present measure of bases of influence. The ranking method used in obtaining the data makes it impossible for all five bases of influence to be correlated in the same direction with any single criterion variable. Thus, it may be that positive correlations with expert and referent influence "caused" negative correlations with other bases of influence, or vice versa. In spite of this limitation, the overall pattern of correlations with bases of influence is of considerable interest.

also parallel, but they tend to be somewhat lower in the case of faculty influence.

DISCUSSION

In general, the results of this study of administration in liberal arts colleges are quite consistent with our hypotheses based upon other organizational studies. It is particularly useful to draw a comparison between the present research and the earlier study of account executives and office managers (Bachman et al., 1966). In both studies satisfaction was positively related to the total amount of influence at all levels. Also in both studies it appeared that satisfaction with an organizational "superior" (office manager or dean) is higher when the superior enjoys a relatively high degree of influence, and when this influence is based upon expertise and personal attractiveness, rather than upon legitimate authority or the use of sanctions. One interesting difference appeared in the rank ordering of bases of the dean's influence. In the study of account executives and office managers, legitimate influence was consistently rated most important (although it was negatively related to satisfaction with the office manager). In the present study, on the other hand, expertise was considered the most prominent basis of the dean's influence, with legitimate authority given a second level of priority. As might be expected, college faculty members are clearly less likely than account executives to see themselves as subordinates in a hierarchy.

Although most faculty members do not consider a legitimate obligation as their most salient reason for doing what the dean wishes, it appears that they do consider this the most important reason why the *dean* complies with *their* desires. Moreover, the greater the relative importance of this basis of faculty influence over the dean, the greater is their satisfaction with him. The explanation for this reversed relationship when "the shoe is on the other foot" is not immediately clear. It should be noted that the two questions are (necessarily) different in wording, and this may contribute to the effect. But it may also be that faculty members really are happy to perceive their rights as being respected by the dean, and at the same time are less satisfied to view their own actions as deferring to the dean's legitimate rights.

In the present study, as in many others, it has been necessary to measure administrative behavior solely in terms of perceptions. Often such a procedure is criticized because what appears to be an "objective" relationship (e.g., deans who rely more upon expert and referent influence are likely to have more satisfied faculty members) may in fact be nothing more than a "phenomenological" effect (e.g., if a faculty member likes a dean, he will then *perceive* the dean as using expert and referent influence—quite apart from the dean's actual behavior). This problem may be clarified by a hypothetical illustration: Professor A at Alpha College and Professor B at Beta College have identical perceptions of their deans' bases of influence. But the mean perceptions of *all* faculty members at Alpha are different from the corresponding means at Beta; the dean at Alpha is viewed *on the average* as making relatively greater use of expert and referent influence. Now if the phenomenological explanation is correct, Professors A and B should not differ in satisfaction with their respective deans. On the other hand, if the use of mean perceptions provides at least some rough indication of objective conditions, and if these conditions influence satisfaction as we have hypothesized, then we should find that Professor A is more satisfied than Professor B. The partial correlations presented in Table 3 provide just this sort of test, and the results clearly support the objective interpretation. When individual perceptions are in effect held constant (thereby removing the possibility of a phenomenological effect), the strong relationships between bases of influence and satisfaction with the dean are scarcely diminished at all. Of course, this type of analysis provides no guarantee that our operations have measured exactly what was intended; but it does indicate that *something* objective has been measured, and that the present findings cannot be dismissed as "mere perceptions."

An adequate empirical foundation for policy recommendations in academic administration awaits much further work. Nevertheless, it may be of interest to summarize a

few tentative implications of the present study in terms of what constitutes an effective dean (*i.e.*, a dean whom the faculty perceives as doing a good job). First, so far as faculty are concerned, the effective dean appears to have a good deal of influence "in determining the policies and practices of the college." Second, there is some support for the view that the faculty see the dean as a colleague or "first among equals" rather than as a hierarchical superior. The effective dean is influential through personal qualities such as expertise and respect, rather than basing his influence upon legitimate authority and the use of punishments or rewards. The collegial relationship should not, however, be construed as minimizing the legitimate authority of the faculty. On the contrary, it would seem that the effective dean is one who obviously appreciates his obligation to respond to faculty suggestions.

One additional implication of the study must be emphasized. It was noted at the outset of this discussion that the present findings are largely similar to those obtained in other organizational studies. This may serve to encourage further efforts applying the findings of organizational research to academia. And it may indicate a common ground of organizational phenomena subject to general theories of administration.

REFERENCES

Bachman, J. G., Smith, C. G., & Slesinger, J. A. Control, performance, and satisfaction: An analysis of structural and individual effects. *Journal of Personality and Social Psychology*, 1966, **4**, 127–136.

French, J. R. P., Jr., & Raven, B. The bases of social power. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory*. Evanston, Ill.: Row, Peterson, 1960. Pp. 607–623.

Likert, R. *New patterns of management*. New York: McGraw-Hill, 1961.

Price, D. C. The influence of the supervisor within his department as a conditioner of the way supervisory practice affect employee attitudes. Unpublished doctoral dissertation, University of Michigan, 1955.

Price, D. C. Influence: A key to effective leadership is the first-line supervisor. *Personnel*, November 1955, **29**(3), 3–11.

Tannenbaum, A. S. Control and effectiveness in a voluntary organization. *American Journal of Sociology*, 1961, **67**, 33–66.

Tannenbaum, A. S. Control in organizations: Individual adjustment and organizational performance. *Administrative Science Quarterly*, 1962, **7**, 236–257.

Tannenbaum, A. S., & Bachman, J. G. Structural versus individual effects. *American Journal of Sociology*, 1964, **69**, 585–595.

(Received March 7, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 1, 62–65

VALIDITY OF THE GUILTY-KNOWLEDGE TECHNIQUE: THE EFFECTS OF MOTIVATION [1]

P. O. DAVIDSON

University of Calgary

48 Ss divided into 12 groups took part in an experimental investigation of detection of a simulated crime. For each of 12 crimes, 3 Ss were highly motivated to commit the crime; 1 succeeded, 1 attempted but failed, and 1 did not make an attempt. The 4th S interrogated for each crime was an innocent control. Motivation for deception (hope of reward) was high ($25–$50) for half of the crimes and low (< $1) for the other half. Detection of deception using a polygraph recording of GSR and the guilty-knowledge technique resulted in correct classification of 92% of "guilty" Ss and 100% of "innocent" Ss. No significant differences were found due to level of motivation for deception. Results were discussed in terms of their relationship to similar studies and actual criminal investigations.

The first recommendation arising out of Orlansky's (1962) review on lie-detection capability was for further investigation of the validity of lie-detection techniques. This view has been reiterated recently in investigations by the Committee on Government Operations (1965, 1966) of the United States House of Representatives concerning the use of polygraphs as "lie detectors" by the Federal Government. A common problem in studying the validity of lie detection, although not commonly recognized by many lie-detector operators, is that the polygraph merely indicates that a physiological response has occurred. The operator must *infer* the lie on the basis of the observed physiological response. In spite of some extravagant claims (cf. Lee, 1953), the inference that autonomic responses can indicate lying must be viewed with extreme skepticism since the physiological responses measured by polygraphs can be produced by a wide variety of stimuli quite incidental to any process of lying. Lykken (1959) has suggested that the detection of guilt by using physiological measurements involves a more reasonable inference than the detection of lying. If a guilty individual is presented with some relevant details about his crime interspersed among similar but irrelevant items, he would be expected to respond differently than an innocent individual who had no "guilty knowledge" about which items were relevant. The assumption here is not that the subject (*S*) will respond in a particular, recognizable way when he is lying, but merely that he will respond differently to relevant details of his crime than to irrelevant details.

Lykken found that he could correctly identify, using a simple objective scoring procedure, all innocent *S*s and 93.9% of his guilty *S*s against a chance expectancy of 50%. He noted, however, that his *S*s were not particularly emotionally involved in their crime, nor were they strongly motivated to try and defeat the test. In a later study he found this guilty-knowledge technique highly resistant to faking (Lykken, 1960). The present study was designed to evaluate the validity of the guilty-knowledge technique under more ego-involving circumstances and varying amounts of motivation.

Method

The *S*s used in this experiment were 48 volunteer college students assigned at random to 12 groups. Three *S*s chosen at random from each group took part in the crime as "hunters"; the fourth *S* did not and he had no knowledge of the nature of the experiment. The "hunter" game was used for enacting each crime since it seemed to be a highly ego-involving game for university students. The three

[1] The author gratefully acknowledges the assistance of Geri Gold and the technical advice of Arthur R. Roberts of the Calgary City Police Department in carrying out this study which was supported in part by Canadian NRC Grant No. APA 213.

hunters from each group were seen individually by the experimenter (*E*) and shown a photograph of a student who was to be their intended victim (the 12 victims were all collaborating psychology students). Basic information concerning the name of the victim, his faculty, and his lecture schedule was supplied. Each hunter was told that the victim would be carrying a sealed envelope which the hunter was to take and examine if he could successfully hunt down the victim and enact his murder. The hunter was informed that the envelopes contained vouchers for prizes ranging in value up to $50 and that he would be permitted to cash the voucher if he was not detected as the criminal after a polygraph examination. The hunter would not know the value of the voucher unless he was successful in enacting the crime. Although the hunters were not aware to it, half of the 12 victims had envelopes with vouchers ranging in value from $25 to $50, while the other half ranged in value from 10¢ to $1. Each hunter was informed that there were other hunters after the same victim and that the first successful hunter would be the one to get the envelope. Each hunter was free to choose his own method, time, and place for enacting the crime; the only restrictions were that he must work alone and tell no one of his plans or success. As an additional incentive to the hunters, prizes were awarded upon completion of the experiment for the most ingeniously devised crimes, whether the hunter was successful in getting the envelope or not.

The hunters were told that in order to add some uncertainty to the success of the crime the victim had been given a random timetable with specified time intervals on it when any crime attempted would not be successful. If an attempt was made during one of these "safe periods" the crime was said to have failed, the envelope was not surrendered, and the hunter was barred from trying again for 24 hr. The hunter would not know until he made an attempt whether it would succeed or not. In fact, no such timetable existed and the first hunter to make an attempt was always told that his victim was in a "safe period" and the attempt had failed. The second hunter to make an attempt was always judged successful and took the envelope.

Each victim had a code number which was posted on a central campus bulletin. This number was to be checked before making any attempt and to be removed as soon as any hunter was successful in obtaining the envelope by enacting the murder of his victim. Thus, if the number was gone, the other hunters would know the game was over and would not make any further attempts on the victim.

In this way, for every victim there were four suspects: one who had committed the crime, one who had tried and failed, one who was motivated to attempt the crime but had no opportunity, and a fourth "suspect" who in fact knew nothing about the crime or the victim and was in no way involved or motivated.

As soon as the second hunter had succeeded in obtaining the envelope the victim supplied *E* with answers to some questions relevant to the crime (e.g., where it occurred, murder weapon used, etc.). Six multiple-choice questions about the crime and the content of the envelope were made up with five alternatives for each question (cf. Lykken, 1959). The first alternate on every question was always a foil and not scored. This was done in order to reduce the possibility of incorrectly scoring orienting responses. The four suspects were then brought in for interrogation by another *E* who had no knowledge about who the criminal was, or which answers to the multiple-choice questions were correct. The *S* was seated in the interrogation room, GSR electrodes were attached to his dominant hand, and a pair of headphones adjusted to his ears. The *E* was located with a Grass Model 7 polygraph in an adjoining room and spoke to *S* via a microphone. Each *S* was told he was to be questioned in relation to a crime. He was to listen to each question consisting of several parts but to make no answer. The standardized list of six questions was then read to *S* and his GSR responses to each of the five alternates were recorded. Heart rate and respiration rate were monitored but not scored.

Scoring of the GSR response after the recording was completed was simple and objective (Lykken, 1959). Scoring was done without prior knowledge of the correct answers in order to prevent any criterion contamination. An *S*'s GSR's to the several alternatives in a given question were ranked in order of amplitude. If his largest response was found to be a relevant alternative he was given a score of 2 on that question. If his second largest response was to the relevant alternative he was given a score of 1. Thus a perfect innocent score was 0 and a perfect guilty score was 12.

RESULTS

Following Lykken's (1959) criteria, all scores of 6 or less were classified "innocent" and all scores over 6 "guilty." All six of the guilty *S*s who were highly motivated to deception because of the high value of their vouchers were correctly classified. Five of the six guilty *S*s who were under low motivation to deception (negligible value of vouchers) were correctly identified. The difference in classification between these two groups was not significant (Fisher exact probability $p = .50$ [Siegel, 1956]). Combining the two groups, 11 of the 12 guilty *S*s were correctly classified as guilty (92%). All 36 of the innocent *S*s (100%) were correctly classified as not guilty. There was a total of 48 interrogations and 47 of these were correctly classified—giving an overall 98% correct classification against a chance level of 25%.

DISCUSSION

There were several reasons for the fairly involved procedure used in this study. Primarily it was designed to simulate, a little closer, actual situations involving real crimes, but at the same time maintain rigorous experimental control. By allowing *S* to choose his own time, place, and method, plus the incentive for ingenious crimes, *S* became much more actively and personally involved than if he merely walked through the motions of a staged crime. Some of the uncertainties involved in real crimes were introduced. Not every crime attempted is automatically successful and there is usually no way of knowing in advance. A criminal usually does not know exactly what his gains will be (if any) until he successfully completes the crime. Furthermore, in order to keep the rewards of his crime he must be able to escape detection. The motivation for escaping detection may be a complex one involving some combination of fear of punishment and hope of reward. Most experimental studies have increased this motivation by threat of punishment (usually electric shock). In the present study, the hope of reward was experimentally varied to see if this had any effect on ability to escape detection in a polygraph examination. It has sometimes been suggested that successful detection rates in actual crimes would be significantly higher than in experimental studies since the tremendous motivation involved in the former would increase the amount of autonomic responsiveness to questions relevant to the crime. While the differences in this study are in this direction, they do not support this hypothesis. They do not clearly refute it either. The small samples of guilty *S*s and the limited rewards combined with no real threats of punishment in the present study require caution in interpreting these results.

One observation did occur which merits further study. A common attitude in the lie-detection field (cf. Lee, 1953) is that the GSR is not useful because it is "too sensitive." Lykken (1960) discounted this objection to the GSR on the basis of his experimental studies with essentially fairly low levels of motivation to deception. In the present study it was observed, particularly in guilty *S*s who possessed $50 vouchers, that there seemed to be a great many spontaneous fluctuations in base-line GSR. Scoring GSR responses to relevant questions must be done very carefully in such cases to prevent the confusion of spontaneous base-line fluctuations from being counted as GSRs. While this was not too difficult in the present experiment, it does indicate caution in uncritical adoption of Lykken's scoring criteria for criminal investigations. Malmo (1957) has suggested that base-line fluctuations can be a useful index of drive level and, consequently, a very high drive level produced in a suspect for any reason (e.g., merely being suspected of a crime) would reduce the reliability of scoring GSR responses. In this sense the "extraordinary sensitivity of the GSR" may not be a "clear virtue" as Lykken (1960) suggests that it is.

Some experimental studies stack the odds for correct detection very much in favor of the polygraph examiner by merely requiring him to separate one guilty suspect from several innocent controls. In criminal investigations the odds are not always so nice. Often there are several suspects who had good motives for committing the crime, some may have been planning it and some may have even attempted it unsuccessfully. But to the extent that these experimental results can be generalized, the guilty-knowledge technique seems to work as well in either situation. The 92% correct classification of guilty *S*s and 100% correct classification of innocent *S*s is essentially the same using the present procedure as the 93.9% correct classification of guilty *S*s and 100% correct classification of innocent *S*s using Lykken's (1959) procedure.

REFERENCES

COMMITTEE ON GOVERNMENT OPERATIONS. (10th report) *Use of polygraphs as "lie detectors" by the federal government.* 89th Congress, 1st Session, House Report No. 198. Washington, D. C., 1965.

COMMITTEE ON GOVERNMENT OPERATIONS. (39th report) *Use of polygraphs as "lie detectors" by the federal government (Part 2).* 89th Congress, 2nd Session, House Report No. 2081. Washington, D.C.: 1966.

LEE, C. D. *The instrumental detection of deception.* Springfield, Ill.: Charles C Thomas, 1953.

LYKKEN, D. T. The GSR in the detection of guilt. *Journal of Applied Psychology*, 1959, **43**, 385–388.

LYKKEN, D. T. The validity of the guilty knowledge technique: The effects of faking. *Journal of Applied Psychology*, 1960, **44**, 258–262.

MALMO, R. B. Anxiety and behavioral arousal. *Psychological Review*, 1957, **64**, 276–287.

ORLANSKY, J. *An assessment of lie detection capability*. Institute of Defence Analyses, Research and Engineering Support Division. Technical Report No. 62–16. Washington, D. C., 1962.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1960.

(Received March 10, 1967)

List of Manuscripts Accepted for the

Journal of Applied Psychology

Goals and Intentions as Mediators of the Effects of Monetary Incentives on Behavior: Edwin A. Locke,* Judith F. Bryan, and Lorne M. Kendall: Department of Psychology, University of Maryland, College Park, Maryland 20740.

Relation between Ethnic Origin and GSR Reactivity in Psychophysiological Detection: Sol Kugelmass* and Israel Lieblich: Department of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel.

Influence of Training, Method, and Relationship on the Halo Effect: Eva Metzger Brown*: Apartment 134, 115–25 Metropolitan Avenue, Kew Gardens, New York 11418.

Driver Gap Acceptance at Intersections: William L. Gibbs*: United States Department of Transportation, Federal Highway Administration, Bureau of Public Roads, Washington, D. C. 20591.

Relationship between Perceptual Style and Driver Reaction to an Emergency Situation: Gerald V. Barrett* and Carl L. Thornton: Director Human Factors Laboratory, Life Sciences Research Department, Goodyear Aerospace Corporation, 1210 Massillon Road, Akron, Ohio 44315.

Environmental Frustration and Creative Problem Solving: Bernard L. Hinton*: Graduate School of Business, Indiana University, Bloomington, Indiana 47401.

Effects of the Type of Stimulus Employed and the Level of Subject Awareness on the Detection of Deception: Richard L. Thackray and Martin T. Orne*: Unit for Experimental Psychiatry, University of Pennsylvania, 111 North 49th Street, Philadelphia, Pennsylvania 19139.

Scales for Rating the Taste of Water: William H. Bruvold*: University of California, School of Public Health, Earl Warren Hall, Berkeley, California 94720.

Occupational Validity of the GATB: Stephen E. Benis*: United States Department of Labor, Bureau of Employment Security, Washington, D. C. 20210.

How to Succeed in Business According to Business Students and Managers: Bernard M. Bass*: 401 Bruce Hall, Graduate School of Business, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

---

* Asterisk indicates author for whom the address is supplied.

Journal of Applied Psychology
1968, Vol. 52, No. 1, 66–70

COMPARISON OF BEHAVIORAL STYLES BETWEEN ENTERING AND GRADUATING STUDENTS IN OFFICER CANDIDATE SCHOOL

PETER B. PETERSEN AND GORDON L. LIPPITT

George Washington University

2 groups of students in an officer candidate school (N = 880) are compared to determine their differences in behavioral styles. The instrument used is the Job Analysis and Interest Measurement. There is a difference at the .01 level of significance in the mean scores between entering (N = 561) and graduating (N = 319) students in 4 prechosen behavioral dimensions. Entering students compared with graduating students place a higher value on approval from others. In the 3 other scales considered, the mean graduating students' scores are higher than the entering students' scores. This indicates that graduating students as compared with entering students tend to be more self-assertive, are more likely to be persuasive leaders, and are more prone to like supervisory activities.

A worthwhile evaluation of training may be accomplished by evaluating groups of trainees before and after they are trained. Schein (1965) favors an evaluation of training which includes an observation of trainees after they have returned to their jobs. A control group may be helpful in determining if the changed behavior is actually caused by training. Prior studies tend to indicate that the change process may be conducted in a variety of ways. Differences in the change process for the most part seem to center on the degree of influence exercised by the person being changed. Schein (1961) visualizes that "coercive persuasion" is employed in some management development programs. He considers that many managers being developed need to be "unfrozen" because they are psychologically unprepared to change certain attitudes in favor of untried new ones. Bennis (1964) suggests that an indoctrination form of change is being used by many schools. While the authors are interested in the change process, primary attention will focus on finding the specific differences between entering and graduating students, and then investigating why these particular dimensions changed. The test instrument used in this study to compare these two groups of students is the Job Analysis and Interest Measurement (JAIM).[1] It has been used for studying more than 30 occupational groups including engineers, lawyers, foreign service officers, ambassadors, judges, social workers, policemen, physicists, and secretaries. Walther (1964) found that this instrument can distinguish among occupational groups and between superior and weak performances within an occupational group. It is a 125-item self-report questionnaire used to measure the personal qualities of the worker (other than aptitudes, training, or knowledge), which have an influence on job success or failure.

This study is a comparison of the average behavioral styles between a group of entering and graduating students in one specific army officer candidate school (OCS). Behavioral styles refer to the consistent ways in which a person organizes his physical, emotional, and energy resources. The specific army OCS used in this study, while anonymous, is representative of nine others. The duration of the training is 23 weeks. The mission of OCS is to develop combat leaders. This training, according to a Department of the Army pamphlet (United States Army, 1966), is designed to place the student under physical, mental, and emotional stress to simulate the stress and fatigue of combat. Within this particular branch being studied, 77% of its newly commissioned officers will be provided by this training source, 22% will be provided

[1] The authors are grateful to Regis H. Walther, author of the test instrument, for his personal assistance in accomplishing much of the work described in this paper.

by Reserve Officer Training Corps training conducted in colleges, and 1% of their officers will be provided by the United States Military Academy at West Point. The high percentage of newly commissioned officers provided by this training source indicates a need for this, and similar, studies.

Problem

The authors were impressed by the fact that some students have a propensity to successfully complete OCS, while others are prepared to resign after they are in the course for only a short time, some for only a few hours. An initial thought was that students with a favorable degree of certain behavioral styles might be more likely to successfully complete OCS. A start in making such an idea fruitful might be a comparison of the behavioral styles between entering and graduating students in OCS. This study will attempt to make such a comparison, pertaining to four specific behavioral style scales.

Groups of students (rather than individuals) will be analyzed. A group of entering students will be compared with a different group of graduating students. Behavioral styles will be analyzed in terms of the significant difference between the mean behavioral styles of the two groups. The hypotheses will reflect directly the purpose of this research, which is to compare four specific behavioral style scales between entering and graduating students in OCS.

There are a number of factors which, in the opinion of the authors, influence the behavioral styles of officer candidate students so that they may become qualified to graduate. Of the many factors, four are considered to be especially pertinent to this study. Their pertinence is judged on the basis of their role in determining the culture that characterizes OCS, the social situation of the student within OCS, and the relationship between the student's social and academic experiences. These factors are: (1) the goal and standards of OCS, (2) the course of instruction, (3) the tactical officer, and (4) the system of voluntary resignation.

It is understood that many other factors may influence their behavioral styles. Some of these other factors are: (1) the prior experience and training of the student, (2) the student's relationship with internal factors within the OCS, (3) the student's relationship with various factors external to the OCS, (4) the role of the student within the formal and informal organization, (5) the student's age and physical fitness, and (6) the changing amounts of power, authority, and influence exercised by the student.

TABLE 1

SUMMARY OF PILOT STUDY RESULTS

Scales	*t*-test results*	Difference between the means of entering students ($N = 56$) and graduating students ($N = 33$)
Perseverance	.82*	.67
Persuasive leadership	3.67**	1.58
Self-assertiveness	3.91**	2.19
Supervisory activities	3.16**	1.32
Approval from others	−6.02**	−2.70

* A positive *t* value indicates graduating students are higher in this particular scale, while a negative *t* value indicates that entering students are higher as pertains to this particular scale.
* $p < .05$ with $t = 1.99$ for $df = 87$.
** $p < .01$ with $t = 2.64$ for $df = 87$.

Pilot Study

In the development of the hypotheses of this study an initial thought was that students high in the JAIM Perseverance scale would have a high propensity to graduate and that those low in the JAIM scale of Perseverance would be likely to fail or resign from the course. A pilot study was conducted in June 1966 in which the average behavioral styles of 56 entering students were compared with the average behavioral styles of 33 graduating students. A summary of the results of this pilot study is shown in Table 1.

The hypotheses of this study are stated below:

1. Hypothesis I: Entering students, compared with graduating students, tend to place a higher value on approval from others; therefore, entering students will score significantly higher on the JAIM scale Approval From Others.

2. Hypothesis II: Compared with entering students, graduating students in OCS tend

to be more self-assertive; therefore, graduating students will score significantly higher on the JAIM scale Self-Assertiveness.

3. Hypothesis III: Compared with entering students, graduating students in OCS are more likely to be persuasive leaders; therefore, graduating students will score significantly higher on the JAIM scale Persuasive Leadership.

4. Hypothesis IV: Compared with entering students, graduating students in OCS are more prone to like supervisory activities; therefore, graduating students will score significantly higher on the JAIM scale Supervisory Activities.

METHOD

This study attempts to compare, systematically, several variables in the behavioral styles between entering and graduating students in OCS. This difference is revealed by comparing the mean behavioral styles of a sample from each of these two populations. The hypotheses of this study are concerned with the differences in four specific behavioral style scales represented in the test instrument used in this study. The instrument, called the Job Analysis and Interest Measurement (JAIM), provides measures for 34 scales relating to attitudes, interpersonal behavior, formal organizational behavior, decision-making behavior, work preferences, values, and academic achievement. The difference in the specific scale means, if statistically significant (determined by using Fisher's t), will form the basis for further evaluation, explanation, and possible implications and speculation.

Entering students sampled were administered the JAIM during their first week of training in OCS, while graduating students were sampled in their twentieth to twenty-second week. Questionnaires were administered during the period from June 21, 1966 through October 21, 1966. The conditions of the testing facilities were reasonably equal in all cases. The data preparation included transferring the student answers from each answer sheet to three data-processing cards. Starting October 7, 1966, an administrative-data questionnaire was also completed by each person administered the JAIM. It is represented by a fourth data-processing card for each student. The analysis of the data was accomplished by an IBM 1620 computer.[2]

The method of evaluating the JAIM results in this study is to determine if there is a significant[3] difference between the mean scores on a specific JAIM scale between entering students ($N = 561$) and graduating students ($N = 319$). The significance in the difference of means is tested by conducting a t test of the two means. In this study a positive t value indicates that graduating students are higher in a particular scale, while a negative t value indicates that entering students are higher as pertains to this particular scale. In the evaluation of the two experimental groups, significance at the .01 confidence level is $p < .01$ with $t = 2.58$ for $df = 878$.

The students in the population sampled were homogeneous as to their age, sex (male), state of health, and general educational level. Some Negro students were present, but the proportion of Negroes in this OCS was less than the proportion of Negroes in the Army enlisted ranks. The results of an administrative-data questionnaire completed by enter-

TABLE 2

GROUP AGGREGATES OF INDIVIDUAL CHARACTERISTICS FOR ENTERING ($N = 151$) AND GRADUATING ($N = 152$) STUDENTS

	Entering students	Graduating students
Method of entering service		
Volunteer	61%	72%
Draft	39%	26%
Other (i.e., National Guard Unit called to active duty)	0%	2%
Students presently in class who have been turned back from another class		
Turned back for academic reasons	1%	7%
Turned back for leadership reasons	1%	13%[a]
Turned back for other reasons (i.e., medical)	0%	4%
Educational level		
High school or equivalent	18%	20%
Some college but did not graduate	54%	47%
Graduated from a junior college	5%	5%
Graduated from college	16%	20%
Some postgraduate work	6%	7%
Master's degree or higher	1%	1%

[a] This may indicate that, for this study, difficulties in meeting leadership-proficiency standards have a higher propensity of being corrected than difficulties in meeting academic-proficiency standards. This is probably indicated by the higher percentage of students graduating who have previously been turned back for leadership reasons than those previously turned back for academic reasons. It may also indicate that judgments on leadership proficiency are more subjective and, therefore, less reliable than judgments on academic proficiency.

[2] Computer operations were conducted by David Reimer, staff associate in the Social Research Group of The George Washington University.

[3] The term "significant" is reserved for the meaning statistically significant.

ing and graduating students are shown in Table 2. This questionnaire was designed to help describe the sample. It reflects the aggregate of individual characteristics for a group containing 151 entering students and a group containing 152 graduating students.

Since 13 entering classes and 10 graduating classes were sampled, the writers felt it advisable to determine the variance among different classes. The statistical test selected for this comparison of classes was the analysis of variance, which provides a method of testing for significant differences between means among and between groups. Each of the four JAIM scales was evaluated in terms of the stability of the entering classes' set of 13 means, and the stability of the graduating classes' set of 10 means.

RESULTS

The results of the analysis-of-variance tests for entering and graduating classes are given in Table 3. Among the entering classes, one significant difference ($p < .05$ with $F = 1.77$ for $df = 12$) was found on the Approval From Others scale. The fact that one out of eight comparisons shows a significant difference among groups indicates that different classes vary more than would be expected due to chance. This should be considered when interpreting the data.

As shown in Table 4, the significance in the difference of means is tested by conducting a t test of the two means. In this analysis a positive t value indicates that graduating students are higher in a particular scale and a negative t value indicates that entering students are higher as pertains to a particular scale. In the analysis of the two experimental groups, a significance of at least the .01 confidence level occurs when t is greater than or equal to 2.58 for 878 degrees of freedom.

TABLE 3

ANALYSIS OF VARIANCE AMONG ENTERING AND GRADUATING CLASSES

JAIM scale	F	p
Approval from others		
Entering classes ($N = 13$)[a]	1.967	.05
Graduating classes ($N = 10$)[b]	1.880	—
Persuasive leadership		
Entering classes[a]	.945	—
Graduating classes[b]	1.552	—
Self-assertiveness		
Entering classes[a]	.798	—
Graduating classes[b]	1.065	—
Supervisory activities		
Entering classes[a]	.908	—
Graduating classes[b]	1.026	—

[a] Entering classes ($N = 13$): $p < .05$ with $F = 1.77$ for $df = 12$; $p < .01$ with $F = 2.19$ for $df = 12$.
[b] Graduating classes ($N = 10$): $p < .05$ with $F = 1.91$ for $df = 9$; $p < .01$ with $F = 2.48$ for $df = 9$.

TABLE 4

SUMMARY OF SCORES FOR THE TWO EXPERIMENTAL GROUPS

JAIM scale	t value[a]	p[b]	Hypothesis affirmed[c]
Entering students higher ($N = 561$):			
Approval from others	−7.40	.01	Yes
Participative leadership	−5.43	.01	Not hypothesized
Graduating students higher ($N = 319$):			
Self-assertiveness	5.61	.01	Yes
Persuasive leadership	4.15	.01	Yes
Supervisory activities	2.58	.01	Yes

[a] A positive t value indicates graduating students are higher in this particular scale, while a negative t value indicates that entering students are higher as pertains to this particular scale.
[b] $p < .01$ with $t = 2.58$ for $df = 878$. $p < .05$ with $t = 1.96$ for $df = 878$.
[c] At the .01 level of significance.

Entering students compared with graduating students tend to place a higher value on approval from others, as determined by the Job Analysis Interest Measurement (JAIM). In the other three scales, the graduating students were higher, which indicated the following:

1. Compared with entering students, graduating students in OCS tend to be more self-assertive, as determined by the JAIM.

2. Compared with entering students, graduating students in OCS are more likely to be persuasive leaders, as determined by the JAIM.

3. Compared with entering students, graduating students in OCS are more prone to like supervisory activities, as determined by the JAIM.

There is a significant difference in the mean JAIM scores between groups of entering and graduating students in the four prechosen comparisons established by the four hy-

potheses. Within the structure of the restraints established by earlier assumptions, the four hypotheses of this study are affirmed and accepted.

The Approval From Others scale measures the degree to which the individual values himself by obtaining the approval of others. The results of this study indicate that the Approval From Others scale, located within the "Value" category of scales, has the highest t value among any of the 34 JAIM scales for this particular research study (t value of -7.40). Graduating students appear to be mission oriented: some express their philosophy as "not overly concerned with what my buddies think; it's the mission that counts." In contrast, entering students tend to "worry too much about what their buddies think." It seems that mission orientation is paramount during training, while the approval from others tends to be secondary.

Studies in social psychology during World War II suggest that the personality of the officer candidate is changed to adapt to new standards. Stouffer, Suchman, DeVinney, Star, and Williams (1949) describe this change as follows:

> The hopeful candidate is now subjected to a nearly catastrophic experience, which breaks down to a large extent his previous personality organization. His previous valuations fail him, and in order to find a basis for self-respect, he must adopt new standards or escape from the field. His high motivation to become an officer usually rules out the latter alternative. At the same time, new, appropriate attitudes are built up and established. The catastrophic experience provides a kind of purgatory, a definite demarcation from the candidate's enlisted incarnation that puts a barrier between the new officer and his enlisted memories. It has some of the characteristics of a conversion experience, or the ordeal of a medieval knight. The effect of this ordeal on the officer candidate is not only to attack his previous personality, but to exert a positive influence in the desired direction.

An interpretation that may be drawn is that students are being molded into the behavioral styles required to become successful junior officers. The graduating student is action oriented, with a greater concern for the mission than for the welfare of his subordinates. However, this does not imply that he is not concerned with their welfare. It does imply that concern for subordinates is second only to the accomplishment of the mission.

Suggestions for Further Research

An evident weakness in this research is that the authors cannot state what portion of the difference in behavioral styles is caused by:

1. Students actually changing their behavioral styles during the time considered;
2. Students with certain behavioral styles resigning from the course and subsequently not being sampled as graduating students;
3. Students with certain behavioral styles failing from the course and subsequently not being sampled as graduating students.

Additional research should devise a system so that individual comparisons may be made instead of comparing the mean of each group.

A study conducted a year or two from now may provide answers as to the stability of the behavioral styles "learned" or "changed" while at OCS. To what extent, if any, the individual reverts toward his original behavioral style offers attentive consideration. A most interesting determination in the opinion of the authors is: To what extent are the behavioral styles of students actually changed?

REFERENCES

BENNIS, W. G., BENNE, K. D., & CHIN, R. *The planning of change*. New York: Holt, Rinehart & Winston, 1964.

SCHEIN, E. H. Management development as a process of influence. *Industrial Management Review*, Massachusetts Institute of Technology, School of Industrial Management, May 1961. Pp. 59–77.

SCHEIN, E. H. *Organizational psychology*. Englewood Cliffs, N. J.: Prentice-Hall, 1965.

STOUFFER, S., SUCHMAN, E. A., DEVINNEY, L. C., STAR, S. A., & WILLIAMS, R. M., JR. *The American soldier*. Princeton: Princeton University Press, 1949.

UNITED STATES ARMY. *The OCS story*. Washington, D. C.: Department of the Army Pamphlet 601–1, June 1, 1966.

WALTHER, R. H. *Manual, job analysis and interest measurement*. Princeton: Educational Testing Service, 1964.

(Received March 16, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 1, 71-80

SOCIAL IMMATURITY, INTELLECTUAL ABILITY, AND ADJUSTIVE BEHAVIOR IN COLLEGE[1]

DAVID KIPNIS

Temple University

It was hypothesized that students classified as socially immature, in comparison to less socially immature students, would be more visible to their instructors because of a higher verbal output in class, would have a higher absentee rate, and would underachieve in academic performance. The Ss were male Temple University students enrolled in 23 sections of an introductory psychology course. Classification of social immaturity was by paper-and-pencil test. In support of the 1st hypothesis it was found that socially immature students were significantly more visible to their instructors by virtue of their classroom behavior. That is, as freshman they were evaluated as more disruptive of classroom proceedings ($p < .05$), and as upperclassmen they were evaluated as contributing more to classroom discussion ($p < .05$). In support of the 2nd hypothesis it was found that significantly more socially immature students were absent from class ($p < .01$). Finally, among students in the top tertile of Scholastic Aptitude Test Scores, socially immature students underachieved in an introductory mathematics course ($p < .01$), but not in an introductory psychology course.

In the last several years there has been increased interest in describing adult personality structure in terms of stages that have their origins in the person's psychosocial development. For example, Loevinger (1966) has described at least seven stages or milestones in the development of the ego; Erikson (1959) has formulated this developmental sequence in terms of stages of ego identity; Sullivan, Grant, and Grant (1957) in terms of maturity levels; and more recently Maslow (1964) has described individuals in terms of levels of synergy. Despite the diversity of terms and stages proposed, there appears to be reasonable agreement about the behaviors of persons classified as at relatively early stages of personality development, that is, socially immature individuals. Among the behaviors prominently associated with social immaturity are impulsivity, a lack of acceptance of conventional norms, and an exploitive mode of interpersonal relations.

There is also evidence that persons classified as socially immature tend to fail in socially defined achievement roles such as are associated with high school (Kipnis, 1965a; Roessel, 1954), college (Gough, 1965; Heilbrun, 1965), or job (Kipnis, 1965b). The evidence also suggests that the relation between social immaturity and underachievement is strongest among more intelligent persons. For example, Roessel (1954) found that among more intelligent students those with high *Pd* scores on the MMPI dropped out of high school more frequently than those with low *Pd* scores, but this relation was not present among less intelligent students. Gough (1965) has consistently found among intellectually gifted persons that the Socialization scale of the California Psychological Inventory is related to school achievement (poorly socialized students underachieved), and Heilbrun (1965) has reported a measure of conformity to social values as predictive of college dropouts among the more intelligent, but not the less intelligent.

While the findings are consistent on the relation between achievement and social immaturity, there is less information available on the nature of the day-to-day activities of socially immature individuals that eventually result in their underachievement. A start toward providing this information was made by the present writer in a taxonomic study of the behaviors and attitudes that distinguish persons classified as socially immature (Kipnis, 1966). Among the findings was that socially immature persons did not

[1] This research was aided by a grant from the trustees of Temple University.

react as frequently as less socially immature persons to embarrassing, shameful, or anxiety-provoking situations. Presumably, this lowered level of reactivity would result in the socially immature person not experiencing anxiety over failure and consequently not being motivated to reduce anxiety by striving for achievement. It was also found that socially immature individuals were less accepting of conventional values relating to obeying rules and regulations.[2] This lack of concern over regulations might further provoke failure through a disregard of institutional rules. The foregoing findings however were based upon the intercorrelation of several questionnaires measuring strength of belief in conventional social norms, reactivity to anxiety, and level of social immaturity. There remains the question of whether these differences in attitudes and beliefs are also manifested behaviorally. The present study reports evidence on this point.

Three studies were done. In the first the prediction was tested that socially immature students would participate more in classroom discussions and be more visible to their instructors by virtue of their classroom behavior. This prediction was based upon a prior finding that socially immature students reported that speaking in public was not likely to make them anxious. Accordingly it was expected that they would talk more in class. The second prediction was that socially immature students would be absent more frequently from class. This prediction was based upon the assumption that the behavior of socially immature students would be less constrained by general university regulations forbidding the cutting of classes. The final prediction was that among more intelligent students, the socially immature student in comparison to the less socially immature student should earn lower grade-point averages (GPAs). This prediction was based upon the previously cited literature on the relation between academic achievement and social immaturity among the more intellectually able population.

[2] Socially immature respondents oriented their values around the acquisition of material things.

PROCEDURE

Measure of Social Immaturity

An index of social immaturity was obtained by adding the unweighted scores from two purported measures of social development—the Socialization scale from the CPI and the Insolence scale (development of this latter scale is given below). The use of this combined score was based upon the goal of improving the measure of immaturity. A subsequent section of this paper will report on the relative effectiveness of the combined scores versus the scores taken singly.

In a sample of 573 male Temple University students, the Insolence scale correlated .45 with the Socialization scale (scoring key reversed so that high scores indicated poorly socialized), suggesting that both scores were measuring aspects of the same personality organization. Both scales contributed relatively equal weight to the combined social immaturity index, as indicated by the fact that the means and standard deviations of both scales were almost exactly equal (Socialization scale: $M = 16.70$, $SD = 5.71$; Insolence scale: $M = 17.31$, $SD = 5.53$).

In the present research, persons with social immaturity index scores in the top third of the Temple University distribution (39 or more) were designated as high in social immaturity (high), persons with middle third scores as middle in social immaturity (middle), and persons with bottom third scores of 29 or less were designated as low in social immaturity (low). It should be noted that students classified as low in social immaturity could probably be further subdivided into several higher stages of development if additional cognitive and attitudinal information were available. In general the scales used appear most sensitive in distinguishing persons fixated at earlier stages of development from persons at all hypothesized latter stages.

Insolence scale. This is a 41-item paper-and-pencil biographical-type inventory. If all items were answered in the keyed direction, it would convey the impression of a physically active, aggressive, pleasure-seeking personality whose childhood was characterized by early independence from family and conflicts with grade-school authorities. The scale is based upon 27 of the 58 items in Torrance and Ziller's (1957) Risk Scale to which have been added 14 additional items describing aspects of childhood behavior. All items were selected on the basis of their relation to school and job performance in the United States Navy.

The usefulness of the Insolence scale as a measure of social immaturity has been explored in a series of studies. In several laboratory studies, persons classified as high in social immaturity on the basis of Insolence scale scores were found to be noncompliant and resistant to an experimenter's influence (Kipnis & Wagner, 1965), and to be nonproductive when working for a leader who was low in status and power (Kipnis & Wagner, 1967), and to be easily aroused to aggressive behavior

(Kipnis, 1966). Among more intelligent patients admitted to a psychiatric ward in a Navy hospital, the scale had a biserial correlation of approximately .60 with psychiatrists' diagnoses of character and personality disorder (Kipnis, 1965b).

In the population at large, the Insolence scale has been found to correlate .41 with the Extroversion scale of the Maudsley Personality Inventory, −.41 with the Achiever Personality scale of Fricke's Opinion, Attitude and Interest Survey, .52 with the Sensation Seeking Scale of Zuckerman et al. (1964), .32 with a measure of lack of emotional responsiveness to shame-provoking situations developed by Lykken (1957), and .29 and .25 with the *Pd* and *Ma* scales of the MMPI, respectively. The scale does not correlate with age or general intelligence as estimated by both the Navy's General Classification Test and the Scholastic Aptitude Test.

Socialization scale. This is a 51-item true-false attitude scale from the CPI.[3] Gough (1965) has described the scale as identifying individuals along a continuum of asocial to social behavior and as able to forecast the likelihood that any person will violate the rules and norms of his culture. Results of a wide variety of studies, summarized in Gough (1965), indicate that persons described as asocial by the scale do indeed exhibit such behaviors as delinquency, underachievement in school, violations of paroles, and show a lack of responsiveness to social reinforcement.

Subjects

The *S*s consisted of male Temple University students enrolled in 23 sections of an introductory psychology course, taught by 13 different instructors. Class size averaged 42 students with approximately half of each class being male students. The specific number of *S*s used varied from study to study, and the exact numbers used will be given when the results from each study are presented. Approximately 49% of the *S*s were freshmen and 51% were upperclassmen (mainly sophomores). The *S*s were tested with the Socialization scale and the Insolence scale during the first week of the term.

Intellectual Ability

Intellectual ability was estimated on the basis of the combined verbal and numerical scores from the Scholastic Aptitude Test (SAT). To detect any possible moderating effects of intelligence, *S*s were divided into thirds on the basis of their combined SAT scores, or at the median of the combined SAT distribution. The number of *S*s available for each study dictated the decision as to whether *S*s were divided into two or three equal groups.

[3] Three items were eliminated as inappropriate for the present population. Permission to use the scale for research purposes was granted by Consulting Psychologists Press, Inc., publishers of the CPI.

Classroom Visibility

Nine instructors teaching 13 sections of introductory psychology were asked to rate all male students in terms of their classroom participation.[4] There were between 20 and 28 male students in each section. The evaluations were made in the ninth and tenth weeks of the school term.

The evaluation forms consisted of three scales. Each scale contained the alphabetical listing of all male students in a section, and the instructor was asked to circle the names of the students on each list that were like the description given at the top of the scale. The scales read as follows:

Classroom visibility. "Some students attract the attention of their instructors because of their classroom behavior. These students may attract notice because they consistently ask questions or make comments about the course material. (Consider in this scale the frequency of classroom participation rather than the quality of the participation.) Other students attract attention because they talk to classmates or they do not appear to be attending to the lecture, or they exhibit other forms of inappropriate classroom behavior. Place a check alongside the names of students whose classroom behavior caused you to notice them."

Facilitative behavior. "Some students contribute to the teaching process by asking reasonable questions or by making constructive comments. Place a check next to the names of the students who most frequently contribute to discussion in your class."

Disruptive behavior. "Some students disrupt classroom proceedings by asking inappropriate or hostile questions, by obvious inattention, or by other kinds of disruptive acts. Place a check alongside the names of students who best fit the above description."

Analysis

The ratings on each scale were analyzed by means of a $3 \times 2 \times 2$ factorial design, with three levels of social immaturity, two levels of intellectual ability (median split on SAT scores), and two levels of students' class (freshman versus all upper classes) as the independent variables. The classification of students in terms of class level was added to test the generally held belief that freshmen participate less in classroom discussions than do upperclassmen. It may be noted that the instructors had no systematic knowledge of students' SAT scores, class level, or level of immaturity scores.

For each *S*, a score was obtained by assigning a weight of 1 if the instructor checked his name on a given scale, and a score of 0 if his name was not checked. These scores were used as the dependent measure in the analysis of variance. A total of 285

[4] Four instructors did not participate: two because they knew the immaturity scores of their students, one because he did not notice any students by name, and one instructor because he did not complete the ratings.

TABLE 1

SUMMARY OF THE ANALYSIS OF VARIANCE: TEACHERS' RATINGS OF CONSTRUCTIVE CLASSROOM BEHAVIOR

Source	*df*	*MS*	*F*
Immaturity (I)	2	.1732	1.47
Class level (CL)	1	1.0890	9.24**
Intellectual ability (A)	1	.6792	5.76*
I × CL	2	.4743	4.02*
I × A	2	.2151	1.82
CL × A	1	.5116	4.34*
I × CL × A	2	.1857	1.57
Error	273	.1179	

* $p < .05$.
** $p < .01$.

*S*s were evaluated. Overall, 23% of the students were nominated as visible to the instructor, 14% were nominated as contributing constructively to classroom discussions, and 9% were nominated as disruptive of classroom proceedings. Two percent of the *S*s were evaluated as both contributing to classroom discussion and as being disruptive of classroom procedures. The number of *S*s in each of the nine cells of the design ranged from 17 to 36; therefore the analysis of variance was based upon an unweighted means solution.

RESULTS

The analysis of the instructors' first ratings on classroom visibility yielded no significant relations between visibility and social immaturity or class level. However, there was a significant main effect for intellectual ability—$F(1/273) = 5.74$, $p < .05$. More intelligent students were rated by their instructors as being more noticeable by virtue of their classroom behavior than were the less intelligent students. An average of 30% of *S*s with above-median SAT scores and 17% of *S*s with below-median SAT scores were rated by their instructors as visible.

The second rating scale requested instructors to nominate students whose classroom behavior was constructive and contributed to the teaching process. The results of the analysis of variance of these ratings of constructive behavior are summarized in Table 1. Table 2 shows the percentage of students nominated by their instructors. It can be seen in Table 1 that both class level and intellectual ability were related to ratings of constructive behavior, at the .01 and .05 levels, respectively. However, there was also a significant interaction between these two variables. Inspection of Table 2 reveals the basis for this interaction. In general, upperclassmen who were above the median in intellectual ability were nominated more frequently as contributing to classroom discussions than were below-the-median upperclassmen or freshmen at all levels of intelligence. The last column in Table 2 summarizes these data.

Table 1 also shows a significant interaction between social immaturity and class level ($p < .05$). Primarily, this significant effect was based upon the fact that among upperclassmen, high immature *S*s were evaluated as contributing significantly more to classroom discussion than were middles or lows. Inspection of Table 2 shows that this effect was limited to more intelligent upperclassmen. Within this more intelligent upperclass grouping, 53% of the highs as compared to 19% of the middles and 22% of the lows were evaluated by their instructors as contributing to the classroom discussion. This latter finding is consistent with Hypothesis 1, which predicted that high immature students would be more visible to their instructors. However, it appears that, in terms of constructive behavior, the hypothesis holds only among more intelligent upperclassmen.

The last rating scale asked instructors to nominate those students in their class whose behavior was disruptive of classroom proceedings. Table 3 summarizes the analysis of variance of these ratings, and Table 4 shows the percentage of students, classified in terms of the independent variables, who were nominated by instructors as being disruptive.

TABLE 2

PERCENTAGE OF SUBJECTS EVALUATED AS CONTRIBUTING TO CLASS DISCUSSION

Class level	Intellectual ability	Level of social immaturity			Average
		High	Middle	Low	
Freshmen	Above median	8%	6%	16%	10%
	Below median	3%	15%	8%	9%
Upper classes	Above median	53%	19%	22%	30%
	Below median	16%	5%	17%	13%

Note.—$N = 285$.

TABLE 3

ANALYSIS OF VARIANCE: INSTRUCTORS' RATINGS OF STUDENTS' DISRUPTIVE BEHAVIOR

Source	*df*	*MS*	*F*
Immaturity (I)	2	.2479	3.17*
Class level (CL)	1	.0023	<1
Intellectual ability (A)	1	.1088	1.39
I × CL	2	.2649	3.39*
I × A	2	.0725	<1
CL × A	1	.0906	1.16
I × CL × A	2	.0804	1.03
Error	273	.0781	

* $p < .05$.

Social immaturity was significantly related to ratings of disruptive behavior ($p < .05$). However it can be seen in Table 3 that this relationship was modified by class level, as indicated by the significant interaction between the two variables. Among freshmen, high immature *S*s were nominated more frequently than middle immature or low immature *S*s as being disruptive. Among upperclassmen on the other hand there was no consistent relation between levels of immaturity and disruptive behavior. The findings are again consistent with the general prediction that high immature *S*s would be more visible to their instructors than middle or low immature *S*s. However, in terms of disruptive behavior, the findings indicate that this prediction only holds among freshmen.

Absenteeism

During the twelfth week of the 16-wk. semester, attendance was taken in all classes. Any person absent on the day attendance was taken was assigned a score of 1 and those present a score of 0. In all, 12 instructors took attendance among 459 students; 21% of the students were absent.

Seven instructors habitually took attendance during the semester, and these instructors had 212 students in 11 class sections. The remaining five instructors had never taken attendance in their classes prior to this time. These instructors had 247 students in 10 class sections. It should be noted that three of the five instructors who did not take attendance usually took attendance during other semesters, but refrained this semester at the request of the present writer. The request not to take attendance arose from the author's interest in testing the effects of environmental restraints (the taking of attendance in this case) upon the conforming behavior of high immature *S*s. A prior study had suggested that high immature *S*s tended to conform to social norms when the environment contained restraints against nonconformity (Kipnis & Wagner, 1967).

Attendance was analyzed by means of a 3 × 2 × 2 × 2 analysis of variance with three levels of social immaturity, two levels of restraint (instructors did or did not take attendance), two levels of class (freshman versus upper classes), and two levels of intellectual ability (above and below the median SAT scores) constituting the independent variables. There was an average of 18 *S*s in each of the 24 cells resulting from the four-way classification. The cell with the smallest entry had 13 *S*s and the cell with the largest entry had 30 *S*s. The dependent measure was a score of 1 or 0, depending upon the absence or presence of the student.

RESULTS

Table 5 presents the summary of the analysis of variance of the attendance data. Three of the independent variables showed significant association with attendance; they were social immaturity ($p < .01$), class level ($p < .05$), and restraint ($p < .001$).

Table 6 gives the average percentage of absent *S*s, classified in terms of social immaturity, class level, and restraint. For ease of presentation, classification of *S*s in terms

TABLE 4

PERCENTAGE OF STUDENTS NOMINATED AS DISRUPTIVE BY THEIR INSTRUCTORS

Class level	Intellectual ability	Level of social immaturity		
		High	Middle	Low
Freshman	Above median	21%	6%	0%
	Below median	17%	4%	5%
Upper classes	Above median	6%	21%	6%
	Below median	8%	5%	0%

Note.—$N = 285$.

TABLE 5

SUMMARY OF THE ANALYSIS OF VARIANCE: STUDENT ABSENCES

Source	*df*	*MS*	*F*
Immaturity (I)	2	.859	5.34**
Class level (CL)	1	.901	5.60*
Ability (A)	1	.033	<1
Restraint (R)	1	2.166	13.45***
I × CL	2	.098	<1
I × A	2	.048	<1
I × R	2	.024	<1
CL × A	1	.093	<1
CL × R	1	.082	<1
A × R	1	.197	1.22
I × CL × A	2	.097	<1
I × CL × R	2	.146	<1
I × A × R	2	.004	<1
CL × A × R	1	.065	<1
I × CL × A × R	2	.005	<1
Error	435	.161	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

of intellectual ability has been omitted, since this variable was not related to absenteeism, either by itself or in interaction with the other three independent variables.

Whether or not instructors took attendance yielded the strongest relation with absentee rate. When the instructors exercised their restraining powers by taking attendance, an average of 13% of the students were absent as compared to 27% when these powers were not used. Further, freshmen had a higher absentee rate (25%) than upperclassmen (17%). Finally, as predicted, an average of 29% of the high immature *S*s, 18% of the middle immature *S*s, and 14% of the low immature *S*s was absent.

It appears then that, without knowledge of actual illness which presumably accounted for at least some of the absentees, such factors as the student's personality, his years of experience as a college student, and the use of restraining power by the instructor were significantly related to the presence of a student in class. In the extreme cells we have 42% of the high immature freshmen who were in classes where attendance was not taken absent, as compared to 2% of the low immature upperclassmen who were in classes where attendance was taken.

Academic Achievement

Final grades in introductory psychology and in an introductory mathematics course were used as the primary criteria of academic achievement. The analysis was limited to 216 freshmen completing their second term at Temple University. This restriction was used in order to control the year in which the *S*s had taken mathematics. The necessity for this control was indicated by a preliminary analysis which found that upperclassmen had significantly higher grades in a given course than did freshmen in the same course, even after controlling for intellectual ability.

Grades were analyzed by means of a 3 × 3 factorial design, with three levels of social immaturity and three levels of intellectual ability, as measured by SAT scores, constituting the independent variables. Final grades in both courses, expressed on a 0 (= F) to 4 (= A) scale, were entered as the dependent measures.

Table 7 presents the results of the analysis of variance of mathematics and psychology grades. There was no significant relation between level of immaturity and the final grade in psychology. There was, however, a significant relation between final mathematics grades and level of immaturity ($p < .05$).

Table 8 shows the average mathematics grades for *S*s classified in terms of immaturity level and intellectual ability. High immature *S*s earned lower mathematics grades than middle or low immature *S*s. While the inter-

TABLE 6

PERCENTAGE OF STUDENTS ABSENT

Class level	Restraint	Level of immaturity		
		High	Middle	Low
Freshman	Attendance taken	24%	14%	11%
	Attendance not taken	42%	37%	21%
Upper classes	Attendance taken	22%	6%	2%
	Attendance not taken	27%	16%	20%

Note.—N = 458.

TABLE 7

SUMMARY OF ANALYSIS OF VARIANCE: MATHEMATICS AND INTRODUCTORY PSYCHOLOGY GRADES IN RELATION TO INTELLECTUAL ABILITY AND SOCIAL IMMATURITY (UPPER FRESHMEN ONLY)

Source	Mathematics			Psychology		
	df	*MS*	*F*	*df*	*MS*	*F*
Immaturity (I)	2	3.77	3.12*	2	1.35	1.32
Intellectual ability (A)	2	1.46	1.20	2	12.15	11.89**
I × A	4	2.15	1.77	4	.08	<1
Error	207	1.21		207	1.02	

* $p < .05$.
** $p < .01$.

action between immaturity level and ability was not statistically significant, the trend appears to be that, as intellectual ability increased, the relation between immaturity and mathematics grades also increased. Since these findings were consistent with prior research on the moderating effects of ability, it was decided to compute separate *F* tests between mathematics grades and level of immaturity within each ability level. These separate tests found that the relation between level of immaturity and mathematics was only significant among *S*s who were in the top third of intellectual ability ($p < .01$). The differences in grades between high, middle, and low immature *S*s were not significant among those students in the middle third or bottom third of intellectual ability.

In addition to grades in mathematics, a tally was made of the number of freshmen either on probation or who had dropped out of Temple University at the end of their freshman year. Those freshmen on probation (cumulative GPA of less than 1.60) or who had dropped from Temple University were given a score of 1 and those freshmen not on probation were given a score of 0. These data were subjected to a two-way analysis of variance and the results of this analysis are summarized in Table 9. An additional eight freshmen who were excluded from the previous grade analysis because of incomplete mathematics or psychology grades were included in the present analysis.

The analysis of variance indicated that level of immaturity and intellectual ability were both related to unsatisfactory academic performance at the .05 and .01 levels, respectively. Table 10 shows the proportion of *S*s classified by immaturity and ability level who were on probation or dropped at the end of their freshman year. It can be observed that at all levels of ability a greater proportion of high immature *S*s than of middle immature or low immature *S*s was on probation or had dropped out of Temple University. Once again, to test for possible

TABLE 8

AVERAGE MATHEMATICS GRADES OF SUBJECTS CLASSIFIED BY LEVEL OF IMMATURITY AND INTELLECTUAL ABILITY

Intellectual ability	Level of immaturity		
	High	Middle	Low
High third	1.46	1.80	2.55
Middle third	1.93	2.05	2.27
Low third	1.77	1.93	1.74

Note.—$N = 216$.

TABLE 9

SUMMARY OF ANALYSIS OF VARIANCE: SUBJECTS ON PROBATION OR DROPPED FROM TEMPLE UNIVERSITY (UPPER FRESHMEN ONLY)

Source	*df*	*MS*	*F*
Immaturity (I)	2	.759	4.03*
Intellectual ability (A)	2	1.586	8.43**
I × A	4	.198	1.05
Error	215	.188	

* $p < .05$.
** $p < .01$.

TABLE 10

PERCENTAGE OF STUDENTS ON PROBATION OR DROPPED

Intellectual ability	Level of immaturity		
	High	Middle	Low
High third	25%	7%	4%
Middle third	33%	36%	4%
Low third	50%	38%	37%

moderating effects of intelligence, separate *F* tests relating immaturity level to academic performance were computed within each ability level. The results indicated that only among *Ss* in the middle third of intellectual ability was the relation between immaturity and grades significant ($p < .05$).

In summary, the findings on academic performance suggest that high immature *Ss* tend to do as well as middle immature or low immature *Ss* in an introductory psychology course, but do considerably poorer than middle immature or low immature *Ss* in mathematics courses. This latter finding holds mainly among *Ss* who were in the top third of intellectual ability. In addition, there was a significant trend for more high immature *Ss* than middle immature or low immature *Ss* to be on probation or to have dropped from Temple University by the end of their freshman year. This trend was most marked among *Ss* in the middle third of intellectual ability.

Comparison of the Validity of the Socialization Scale and the Insolence Scale

As a general statement the combined Socialization and Insolence scales yielded slightly stronger relations with the dependent variables studied here than did either scale by itself. Considering each dependent measure in turn, neither scale was related to the instructors' first ratings of general visibility. The Socialization scale was significantly related to evaluations of facilitative classroom behavior ($p < .05$), while the Insolence scale was not. The Insolence scale was significantly related to evaluations of disruptive behavior ($p < .05$), while the Socialization scale was not significantly related to this evaluation. In terms of relationships with absenteeism, the Insolence scale was significantly related to absenteeism ($p < .05$) and the Socialization scale just missed statistical significance ($p < .08$). Both scales were equally related to mathematics grades.

DISCUSSION

It was originally proposed that high immature students would be more visible in class than middle immature or low immature students because they would experience less anxiety over speaking in public and hence should speak more. It was found that as freshmen, high immature students were more visible to their instructors because of their disruptive behavior. As upperclassmen, high immature students were more visible to their instructors because of their facilitative behaviors. While these findings are consistent with the hypothesis of a higher verbal output for high immature students, the changes in evaluation of this output by instructors from freshmen to upper-class years also suggests that some form of social learning may be occurring during this period; that is, high immature students may be learning to modify their verbal output so that what they say is more acceptable to their instructors. Another obvious possibility, of course, is that those high immature students who were rated as disruptive in their freshman year were dropped from Temple University and different high immature students were evaluated as facilitative in subsequent years. Unfortunately, there are no data on the classroom behavior of the same students over a period of several years. Such information would be necessary to clarify the basis for the shift in evaluations (i.e., a longitudinal study).

The second expectation of the present study was that high immature students would be less conforming to general university regulations against "cutting" classes than would middle immature or low immature students. This prediction was based upon previous findings that high immature students had lower internalized restraints against violating rules and regulations than did middle immature or low immature students. Lacking such internalized restraints, it was

expected that high immature students would be more easily instigated by friends or their own impulses to "cut" classes. The findings supported the hypothesis in that significantly more high immature than middle or low immature students were absent from class. It was also found that when instructors enforced university regulations against absenteeism by taking attendance, there was a significant reduction in the number of all students who were absent. This latter finding can be interpreted as indicating that if the environment is perceived to contain punitive restraints against nonconformity, acts of nonconformity will decrease. This interpretation is consistent with more general studies of coercive power (French, Morrison, & Levinger, 1960; French & Raven, 1959; Zipf, 1960) in which it has been found that reliance by authority figures upon threats of punishment induced higher rates of conformity than when threats were not relied upon. In essence, the taking of attendance by instructors represents an exercise of threatening power, since university regulations proscribe punishments for excessive numbers of unexcused absences. And to the extent attendance is taken, the probabilities are higher that students who are continually absent will be reported to the administration.

As a final point it should be noted that the findings on absenteeism among high immature students are not consistent with a hypothesis of Heilbrun's (1965) that regimentation imposed upon students by universities contributed to the academic failures of bright, poorly socialized students. Heilbrun, of course, was referring to a general system of social and academic regimentation imposed upon students. Nevertheless, to the extent that taking attendance imposes restraints upon students, from Heilbrun's hypothesis one would expect increased resistance and increased nonconformity among bright, poorly socialized students in classes where attendance was taken. Actually the present study found a rise in conforming behavior of presumably an adjustive kind among poorly socialized students when restraints were imposed. While there is no long-term evidence on whether this kind of restraint eventually frustrated bright, poorly socialized students and caused them to drop out of school, an analysis of a subsample of second-term, bright (upper-third SAT scores) freshmen's final psychology grades yielded no differences in grades of high immature students in classes that did or did not take attendance.

Turning next to academic performance, the most interesting finding was that among students with high third scores in intellectual ability, high immature students did as well as middle immature or low immature students in a course in introductory psychology, but not as well in a course in introductory mathematics. One likely explanation for these differences in achievement appears related to the task demands imposed upon students by the mathematics and psychology courses. The mathematics course requires constant class attendance and home study, including the solving of assigned problems, in order to follow the developing logic of the course. It is generally agreed that once a student falls behind in his mathematics assignments, it becomes difficult for him to catch up with the class. The psychology course, on the other hand, can be mastered by a reading of the textbook at reasonable intervals. With few exceptions, the quizzes and final test in psychology are taken directly from the textbook. As a result, class notes and attendance do not appear crucial for obtaining a satisfactory psychology grade. Furthermore, the interdependence between various sections of the psychology course is only slight, so that reading the section of the textbook on personality theory, for instance, does not presuppose much knowledge of learning theory which had been taught earlier in the term.

The task requirements of mathematics then appear to require more sustained effort by the student and fewer absences from class than do the task requirements of psychology. Such requirements seem less consonant with the behaviors typically shown by high immature students. While no data were available on this point, the author would speculate that high immature students' underachievement in mathematics resulted from their spending less time in study outside of the classroom and being absent from class more often than middle immature or low immature

students. More generally, these findings point to the possibility that high immature individuals will underachieve in courses in which (*a*) the units of information are highly interdependent so that progress in the latter stages depends upon having learned the earlier stages of information, and (*b*) the course requires sustained effort by the student both in class and in doing homework.

REFERENCES

ERIKSON, E. H. Identity and the life cycle: Selected papers. *Psychological Issues,* 1959, **1**, 1–171.

FRENCH, J. R. P., MORRISON, W. H., & LEVINGER, G. Coercive power and forces affecting conformity. *Journal of Abnormal and Social Psychology,* 1960, **61**, 93–101.

FRENCH, J. R. P., & RAVEN, B. The bases of social power. In D. Cartwright (Ed.), *Studies in social power.* Ann Arbor, Mich.: Research Center for Group Dynamics, 1959. Pp. 150–167.

GOUGH, H. G. Conceptual analysis of psychological test scores and other diagnostic variables. *Journal of Abnormal Psychology,* 1965, **70**, 264–302.

HEILBRUN, A. B. Personality factors in college dropouts. *Journal of Applied Psychology,* 1965, **49**, 1–7.

KIPNIS, D. Intelligence as a modifier of the behavior of character disorders. *Journal of Applied Psychology,* 1965, **49**, 237–242. (a)

KIPNIS, D. The relation between persistence, insolence and performance as a function of general ability. *Educational and Psychological Measurement,* 1965, **25**, 95–110. (b)

KIPNIS, D. Behavioral correlates of character maturity. *American Psychologist,* 1966, **21**, 628. (Abstract)

KIPNIS, D., & WAGNER, C. The interaction of personality and intelligence in task performance. *Educational and Psychological Measurement,* 1965, **25**, 731–744.

KIPNIS, D., & WAGNER, C. Character structure and response to leadership power. *Journal of Experimental Research in Personality,* 1967, **2**, 16–24.

LOEVINGER, J. The meaning and measurement of ego development. *American Psychologist,* 1966, **21**, 195–206.

LYKKEN, D. T. A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology,* 1957, **55**, 6–10.

MASLOW, A. H. Synergy in the society and in the individual. *Journal of Individual Psychology,* 1964, **20**, 153–164.

ROESSEL, F. P. MMPI results for high school dropouts and graduates. *Dissertation Abstracts,* 1954, **14**, 942–943.

SULLIVAN, C., GRANT, M., & GRANT, J. The development of interpersonal maturity. *Psychiatry,* 1957, **20**, 373–385.

TORRANCE, E. P., & ZILLER, R. C. Risk and life experience: Development of a scale for measuring risk-taking tendencies. USAF WADC Tech. Note, 1957. No. 57–23.

ZIPF, S. G. Resistance and conformity under reward and punishment. *Journal of Abnormal and Social Psychology,* 1960, **61**, 102–109.

ZUCKERMAN, M., KOLIN, L. P., & ZOOB, I. Development of a sensation-seeking scale. *Journal of Consulting Psychology,* 1964, **28**, 477–482.

(Received March 22, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 1, 81–88

INTERFACE BETWEEN PERSONNEL AND ORGANIZATIONAL PSYCHOLOGY

BERNARD M. BASS

Management Research Center, Graduate School of Business, University of Pittsburgh

The 1st part of this report delineates the importance of organizational considerations to the personnel psychologist as he concentrates on recruiting, selection, training, job design, and so forth. The 2nd part deals with the converse: the importance of personnel considerations to the organizational psychologist as he deals with problems of morale, supervision, teamwork, organizational design, and conflict resolution.

Among the Inca farmers, work was regarded as an end in itself, a ritualistic form of religious worship. But work that was not ceremonial lacked sense and meaning. Work and worship were inseparable (Von Hagen, 1957). Nevertheless, more often in nonindustrialized societies, work was done merely to subsist. Better yet, it was delegated to women or slaves. The Tibetan concentrated on his prayer wheel; the Talmudist on studying the Law; the Balinese on art; the Athenian on citizenship; and the Iroquois on military honors. Success, well-being, security, and prestige depended not on one's job but on contemplative zeal, scholarship, artistic efforts, oratory, or bravery in battle.

On the other hand, in modern industrialized society, as among the Incas, work is becoming the center of life. Twenty years may be spent preparing to enter the world of work. As much as half of each working day will be spent at work or commuting to and from it. Success in life, sense of well-being, sense of accomplishment, security, and prestige all will be tied to what work is done. Likewise, work will determine one's standard of living and economic well-being. Therefore it should come as no surprise to find that in our "work-intoxicated" society there are few considerations about employment recruiting, testing, hiring, training, job design, and other personnel problems that can be divorced from their setting within a larger context of social forces, organization, and society.

More generally, as Porter (1966) noted in his recent annual review chapter on personnel management, few issues in personnel psychology can be completely examined without attention to what usually are thought to be problems of organizational psychology. It may be profitable at this time to look at some of the more lively issues that lie at the interface between the two fields. Let us examine some organizational questions that appear when dealing with what usually are the concern of technicians in personnel psychology. To do this, I will focus briefly on recruiting, the application blank, testing, interviewing, training, and job design. Following this I will reverse the examination and look briefly at some significant personnel questions, questions of individual differences, which usually are the concern of organizational technicians.

Personnel Issues Involving Organizational Questions

Recruiting

In one large sales organization, approximately 3% of those applicants for jobs as salesmen were hired if they had been attracted to apply for the jobs through newspaper ads. On the other hand, 14% of those who had made application because of friends' suggestions or who had been recommended to the sales agency by friends were hired. Unfortunately, analyses of the merit ratings a year afterwards of those hired found that those who had been recruited through newspaper ads were in the fiftieth percentile in merit while those who had been introduced to the organization through friends were likely to be found below the twentieth percentile in merit. What had been revealed was quite simple to a technician. Different selection ratios were operating.

Those recruited by newspaper advertisements were screened more carefully; proportionately fewer were hired. Those recommended by friends were screened less thoroughly; proportionately more were hired. The same phenomenon was seen quite differently by the social psychologist. For him, interviewers were showing too much acceptance of applicants simply because they had been sent for consideration for employment by friends.

There has been surprisingly little research on the social psychology of applicant sources. Yet, personnel technicians are aware of the overriding importance to the utility of a selection program of the selection ratio as well as the importance of the quality and quantity of recruits who apply for testing. There is little systematic evidence of the many biases involved in decisions about where to recruit and whom to recruit for testing. For example, in academia, we seldom try for a low selection ratio. Rather, we solicit applications for academic jobs from only the most prestigeful sources, neglecting the utilities of uncovering a larger number of recruits. The problem is complicated further when the recruits gravitate toward or away from certain occupations and organizations because of extraneous factors. Consider how difficult it is to hire Negro engineers when the problem is compounded by the difficulties in encouraging Negroes to prepare for careers in engineering, since traditionally it has been so difficult for them to enter such a career.

Application Blank

In a sales selection program, serendipity revealed two *dynamic items* on an application blank (dynamic application questions concern the interplay between applicant and organization). When combined, these two items yielded a phenomenal validity of .70 against subsequent merit ratings. Hindsight made it relatively easy to understand why these two items were such powerful predictors of subsequent job performance. These items were: "Is it all right to contact your boss?" and "How soon can you start work?" Applicants who subsequently were meritorious on the job gave an unqualified "yes" in answer to the first question. It was perfectly all right to contact their boss. Applicants who did not turn out well on the job either said "no" or "yes, but. . . ." Those who turned into poor performers were ready to start work immediately, while those who insisted on at least 2-weeks notice were likely to be better salesmen, subsequently. Probably, what was being assessed by these dynamic items was the sense of responsibility and sense of security of applicants.

The increasing attention we are giving to biographical information blanks points to a likelihood of focusing more fully on the social and organizational history of the appliant, for so much of what is contained in biographical information is of this form. Consider some of the factors which emerged in Cassen's (1966) factor analyses of biographical information blank items in three cultures: upward mobility, family attitudes, achievement by conforming, interpersonal relations, etc.

Testing

There has been a merger of interest of social and personnel psychologists concerning attitudes toward work. There seems to be no simple response to the question "Why men work?" Much research suggests quite strongly that what may be rewarding for the task-oriented worker may be punishing to the interaction-oriented worker (Bass, 1967b). Noncognitive dimensions of particular importance for selection and placement testing lie close to work itself or to organizational considerations. Heretofore, to a considerable degree, personality and motivation assessment have developed within counseling and clinical psychology and have been lifted bodily for application to industry. There has been more concern about neuroticism than there has been for need for achievement, yet there is likely to be more payoff from concentrating on the latter. Even in the case of hospitalized mental patients, it seems more prognostic in vocational rehabilitation efforts for them to know about how task oriented they are than how neurotic they are (Distefano & Pryer, 1964).

There has been little further exploration so far of some intriguing evidence that ability and aptitude tests tend to be more predictive of early success in organizations while interest and personality factors tend only to become useful for prediction of criteria where the

criteria are measured much later in the applicant's subsequent career. This proposition as well as others which are being generated about the dynamic nature of criteria cannot help but result in a more sophisticated approach to forecasting using psychometric measurements.

Early in the history of industrial psychology, the miniature test was devised for selection purposes. Replicas of the real-life job were constructed and performance evaluated in these simulated situations. Now we see much more use of such simulations of organizations for assessment purposes. Such simulations can be accomplished through the "in-basket" technique, leaderless discussions, business games, organizational design problems, and the like. To date, there has been relatively little exploration of the possibilities of creating for assessment purposes relatively simple games, each representing important organizational dilemmas with which to confront applicants who are seeking careers in those organizations.

Culture fair testing is still another critical issue lying between personnel and social psychology. Is culture fair testing even theoretically possible if past behavior is the best predictor of future behavior? Is it possible to rule out cultural influences in testing without eliminating the validity of the tests? To what extent does previous organizational history play a role in test responses? How different is the applicant of today who grows up in a school system where the administration of objective tests is an almost weekly routine from the applicant of the last generation for whom objective testing was more of a novelty?

As long as there is a basic mistrust between the applicant and the organization, we will live in a situation where the sensible applicant attempts to distort and fake his results on noncognitive tests. In what ways could sufficient trust be promoted between applicant and organization to increase the applicant's desire to give honest responses rather than socially desirable ones, or responses which he judges to be ideal or like those of his boss or like those that would land him the job?

Still another issue in testing which may involve public and organizational policy questions deals with the selection ratio and how it is set.

In one northeastern state, the Human Relations Commission suggests that cut-off scores should be no higher than necessary to screen out unqualified applicants. The company which exploits the selection ratio by recruiting more applicants and setting higher cut-off scores does so at the expense of discouraging applicants who score reasonably well on the tests in question. Which is more important: more efficient use of screening devices by the company or avoiding the discouragement of average applicants?

Interviewing

Studies of the bases upon which interviewers make their decisions are beginning to appear more frequently. They may point ultimately to an interview which is a situational test, a replica of an organizational problem where performance can be evaluated completely objectively, primarily by attention to the process by which the interviewee copes with the interviewer. This most promising approach to interviewing treats it primarily as "the social psychology of two-person interaction." Thus Yonge (1956) permitted the interviewer considerable latitude as far as the content of the interview was concerned, but had him rate the interview process itself, specifically to assess the social skill and motivation demonstrated during this process by the interviewee.

The interview presents a particular challenge to the personnel and social psychologist, for no matter what can or cannot be demonstrated about its utility as an assessment device, it is almost universal in use for selection purposes. Above and beyond consideration of the interview as needed for transmittal of information about the job and organization to the interviewee and for helping the interviewee develop realistic expectations about the job for which he is applying, the typical employer still feels the need for face-to-face contact with the applicant before making the employment decision. Yet, it may be that even where the interviewer has been found to add validity to the assessment predictions above and beyond what could be done by statistical integration alone, such

results may be handled more adequately in the future by better statistics. For instance, more attention to modifier variables may be required. Actually, modifiers may be merely symptoms of curvilinearity in the predictors. Under such circumstances, a Bayesian solution will always yield more accurate decisions, and in the last analysis will match the regression decisions when the data is linear (Clampett, 1966). At best, Sawyer's (1966) comprehensive review of research on methods for combining assessments strongly indicates that predictive efficiency is increased when interviewer judgments are mechanically combined or synthesized with test data to yield final predictions in contrast to following the usual practice of having interviewers make clinical judgments about the job applicant from his test scores and interview performance.

The problem and its solutions therefore may be stated as follows: no matter what, interviews will be held with applicants. To insure that the interviews do as little damage as possible to the accuracy of the entire selection process, it may be best to increase the number of recruits who are tested, thereby reducing the selection ratio and at the same time increasing the likelihood that all those who passed the first test screen, and as a consequence receive interviews, are all fairly good prospects for the job in question. At the same time, the interviewer's judgments should be treated like additional measurements to be combined statistically with test data available on the applicant to provide a final prediction for him. This required change in practice is an organizational question since the interviewer's role is reduced in importance. Resistance is likely particularly if the interviewer has considerable status or experience but in its resolution may lie the major portion of the utility of a selection program.

Training

Commenting on the fads and fashions in training is like commenting about the weather and what can be done about it. Nevertheless, organizational matters are vital in the decisions of training programs to be introduced, continued, or abandoned. Let me enumerate some well-known training problems at the interface of personnel and organizational psychology.

1. Trainers have discovered to their chagrin that it is one thing to maintain an effective program which meets its training objectives, but it is another to create a training program which both meets its objectives and obtains the approval of the trainees. Unfortunately, evaluation and continuation of the program are more likely to rest upon the latter rather than the former.

2. Familiar also is the organizational problem associated with the returning trainee. Fleishman (1953) provided sufficient evidence that whether or not the effects of the training program made their appearance on the job 6 months after training depended primarily on the attitudes of the boss and the climate of the organization to which the trainee returned.

3. The professional and the manager need to see themselves engaged in "life-long learning" to avoid obsolescence. The organization can help considerably to maintain expectations that its members must continue to keep up with new developments.

4. To a considerable extent much more attention may need to be paid to a basic conflict between management trainer and management trainee. Managers see themselves performing a complex art whereas one may be trying to teach them simple science. A recent unpublished experiment by Alex Bavelas is relevant here and illustrates the tremendous difficulties involved in the training of people under these circumstances. One subject learns to discriminate the slide photos of healthy and sick cellular tissue through appropriate reinforcement of his responses. A second subject receives the same reinforcement schedule regardless of what responses he makes. The first subject forms a few simple hypotheses about what differentiates photos of a healthy and a sick cell; the second subject forms a complex set of hypotheses since his reinforcements have not been associated in any simple way with his different responses. Unfortunately, when the two subjects discuss the matter, the one with the complex art of judgment is more confident, more resistant to change, and less readily in-

fluenced than the other subject with the simple hypotheses.

Job Design

As the organization introduces automation, there is a flattening of the distribution of skill demands on newly created jobs and the remaining old ones. More routine jobs as well as more skilled jobs emerge while those at intermediate levels of skill demand are abolished. One can only vaguely foresee the creation of less-fluid worker castes, the unskilled, and the elite, unless some effort is made to intervene with designs for more functional intermediate-level jobs. For instance, electronic data processing (EDP) has created the elite programming job. It has also created the unskilled job of pressing bent IBM cards which are returned by mail.

At the level of the manager, there seem to be two diametrically opposed predictions about how EDP is changing the organization of the future and the jobs within it. On the one hand Leavitt and Whisler (1958) see greater centralization decreasing responsibilities accorded middle managers; others see the reverse. Selection and training of future middle managers need to take account of EDP, but how to do so in the face of the uncertain effects of EDP is a dilemma we face at this time. Individual differences play an important role here. For instance, younger men at the same level in the organization see that their own jobs are changing as a consequence of EDP much more so than do older men at the same level of management (Vaughan & Porat, 1967).

Compensation

For too long the questions concerning equitable compensation have been left in the hands of organizational rationalists. We are just beginning to see the full extent of individual differences in attitudes toward compensation. For instance, in a simulation we use for training managers, the managers must award salary increases to 10 engineers each of whom differs in merit and in other attributes. Each has a job somewhat different than those of the others who are to be assigned salary increases. Wide differences appear in the average increases recommended. North Europeans seem to feel that 4–6% is equitable, while South Europeans may push for as high as 36%. Obviously, differences in the rate of inflation in the economies of the different countries in question are of consequence. But just as important are the attitudes toward technicians and engineers that vary greatly from one country to the next. In this same exercise, American, British, and Irish managers generally seem unmoved by the possibility of losing good men to competition and do not feel that counteroffers should be the basis for salary increases. In fact, they sometimes tend to punish men who receive offers from other firms. Contrarily, Flemings, Norwegians, Italians, Indians, and Latin-Americans tend to be more prone to award relatively large increases to men with counteroffers. Even among managers from the same firm, we can see some opting for extremely complex differentials while others insist that regardless of merit or job or seniority, all engineers ought to receive the exact salary increase (Bass, 1967a).

Criteria of Individual Performance

Concern for an employee's performance in a firm must be viewed in terms of his firm's objectives. For example, Smith sells more merchandise faster than Jones, but Smith's customers complain more about their purchases. Smith brings in more new, nonrepeat business, but Jones has more steady, satisfied customers. Who is the better salesman? The answer depends on the firm's goals on what it values most. Smith is the better salesman if the firm is concerned most about its current share of the market. Jones is the better salesman if the firm is concerned about its long-term standing in the market.

We must understand something about how to assess organizational worth if we are to appreciate how individual personnel differ in their contributions to it. Individual differences in the proficiency of executives, for instance, can only be made meaningful if we understand the purposes of their jobs. This in turn entails determining the purposes of the department's divisions and ultimately the purpose of their enterprise (Fiske, 1951). Assessing the adequacy of the performance of a manager can be tricky business indeed. As

Shartle (1956) pointed out, one may discover 10 years too late that an executive who contributed the most to the firm had been discharged.

ORGANIZATIONAL ISSUES INVOLVING PERSONNEL QUESTIONS

We have seen that the problems of recruiting, application forms, testing, interviewing, training, job design, and compensation, all ordinarily the primary interest of the personnel psychologist, are likely to contain social issues of interest to organizational psychologists, as well. Now let us reverse roles and look at some problems which usually are the concern of organizational technicians which nevertheless contain significant personnel questions as well: supervision, communications, conflict resolution, team composition, and organization design.

Supervision

For 25 years, there has been major interest by social psychologists in the utility of democratic or permissive rather than directive supervision. But evidence is continuing to accumulate that what type of supervision works best is often a matter of individual and cultural differences. In a number of studies, more directive approaches are favored by subordinates. Indeed, much direction is more often expected in many locales. For instance, in a recent pilot study completed in Spain, those who acted as subordinates were much more favorably disposed toward supervisors who attempted to persuade them rather than who attempted to share the decisions with them, for, they said, a supervisor who does not try to influence his subordinates fails to accord the subordinates the dignity to which they are accustomed by showing that he really does not have sufficient interest to bringing them over to his position. Numerous studies in the United States suggest that those who are highly authoritarian prefer in turn to submit to authority rather than to operate in an environment with opportunities for sharing decisions with their superiors. In turn, they expect to make decisions for their own subordinates (Bass, 1965).

Communications

In a similar vein, two-way communication is thought more effective as well as more satisfying than one-way communication. Yet, a group of Japanese frustrated an American management trainer who was trying to demonstrate the differences and relatively greater values of two-way communication. Among the Japanese, with whom he was working, the ease of communicating a pattern of rectangles one-way was somewhat greater for the task which was imposed, and as a consequence there was considerable confusion on the part of the students who were ostensibly being taught and shown the value of two-way over one-way communication. We have had similar experiences with engineering students who if they are communicating the pattern of rectangles to others with a great deal of background in mechanical drawing and blueprint reading, may communicate faster and more effectively one-way than two-way. Given individuals with common codes, one-way communication can be more effective and more satisfying than two-way communication.

Conflict Resolution

Some zero-sum games where one party wins only under the condition that the other party loses are situations where conflict resolution is impossible for almost anybody. In some mixed-motive games, resolution will or will not occur, depending on who is playing the game. If one is able to compete with a generous opponent, it is possible to fleece him. If one himself is nurturant, speedy and highly satisfying resolutions can be achieved (Loomis, 1959). There would seem to be considerable payoff in the study of the interaction of person and structural conditions in various types of negotiating situations. Thus, Bass (1966) found that task-oriented negotiators were most likely to achieve settlements of high quality in contrast to settlements reached by less task-oriented negotiators.

Team Composition

We are beginning to understand the personal ingredients in the assembly of members of a group required to achieve particular

outcomes. For example, evidence accumulated so far suggests if we put together a number of highly task-oriented members, we are likely to generate a great deal of socio-emotional conflict, although at the same time the chances are reasonably good that plenty of work will get done. At the same time if we mix together a group of primarily interaction-oriented personnel, the likelihood is that the group will be most satisfying to the members, there will be much play, but relatively little work will be accomplished (Bass, 1967b).

Organizational Design

To conclude, let us look at what rational organizational designers would suggest as inviolate principles of organizational design and at the same time note the very opposite points of view voiced by behavioralists.

The rationalists would say that an organization should be designed so that someone is responsible for supervising all essential activities. The behavioralists would be primarily concerned with creating structures where leadership was shared.

The rationalists would argue that responsibility for specific acts should not be duplicated or overlapping. The behavioralists would feel that overlapping had much merit, providing opportunities for cross-training, backup, and increasing the reliability of the system of interacting workers.

The rationalists would argue for job simplification; the behavioralists would push for the reverse, job enlargement.

The rationalists would insist that responsibilities should be written, clear, and understood by job occupants. The behavioralists would argue that each person brings to a job somewhat different potentials and should be given freedom and flexibility to develop his own particular way of doing the job to make the most of the situation in which he finds himself.

The rationalists would argue that authority to make decisions should be commensurate with responsibility for those decisions. The behavioralists would state that authority cannot be assigned but rather goes with ability and esteem as an individual among associates. Furthermore, they would note that the rationalists' principle of authority primarily makes it easy for some people to shirk responsibility and also increases staff-line conflict.

The rationalists would want authority to be delegated so that decisions take place as close as possible to the point of action. The behavioralists would insist that the rationalists do not mean what they say, for to accomplish this goal they would accept the behavioralists' position that individuals should be responsible for decisions which affect themselves, whereas, in fact, the rationalists design organizations so that such decisions are lodged with the superiors of the individuals who must execute those decisions.

The rationalists have fixed notions about span of control; the behavioralists feel that the span of control creates as many problems as it solves. For example, the smaller the span of control in an organization the taller will be its structure resulting in greater separation of the top of the organization from the bottom. More possibilities will arise for filtering communications that must be transmitted through the organization. There will be greater differences between the goals of those at the top from those at the bottom of the hierarchy. Rather than concern themselves about span of control, the behavioralists would be much more interested in the opportunities for feedback between subordinate and superior; span of control could be much greater where feedback was accurate and easy.

The rationalists insist on a chain of command; the behavioralists say the chain of command is often a fiction interfering with the flow of communication that is needed by the organization, a flow which may be horizontal, diagonal, and in forms quite different than assumed by the rationalists. For example, the relation between the foreman and an assembly-line worker has been likened to that between a travel consultant and his client. The "real boss" in the situation is the machine (Bass, 1965).

Optimum organizational designs are likely to be found somewhere between rationalist emphasis on predictability and accountability through impersonal structure and behavioralist concern for interpersonal trust and interpersonal confidence as the bases of organizational stability and growth. Again, one can

only agree that organizational design must attend to what has been learned about intrapersonal as well as interpersonal dynamics. Where the optimum lies between the rationalist's and the behavioralist's positions depends on the capabilities, training, and involvement of the personnel in the system. The fully programmed rationalist's organization is closer to the optimum when personnel capability, training, and involvement are low (and the rationalist's organization is likely to keep personnel involvement low). When personnel potential is high, the behavioralist's model becomes more feasible.

The personnel specialist no longer can hide behind his validity coefficients, test blanks, and training manuals. In turn, the organization scientist cannot remain assured of the generality of results from his surveys, field studies, and rational analyses. Models for organizational and personnel research need to take into account sources of variance due to jobs and organizational environment, sources due to individual characteristics, and, most important of all, sources due to the peculiar interactions of individuals and environments.

REFERENCES

BASS, B. M. *Organizational psychology*. Boston: Allyn & Bacon, 1965.

BASS, B. M. Effects on the subsequent performance of negotiators of studying issues or planning strategies alone or in groups. *Psychological Monographs*, 1966, **80**(6, Whole No. 614).

BASS, B. M. Combining management training and research. *Training Directors Journal*, 1967, **21**(4), 2–7. (a)

BASS, B. M. Social behavior and the orientation inventory: A review. *Psychological Bulletin*, 1967, **68**, 260–292. (b)

CASSENS, F. W. Cross-cultural dimensions of executive life history antecedents. Unpublished doctoral dissertation, Louisiana State University, 1966.

CLAMPETT, H. A., JR. Psychological predictions based on Bayesian probabilities. Technical Report No. 11, March 1966, University of Pittsburgh, Contract Nonr 624(14).

DISTEFANO, M. K., JR., & PRYER, M. W. Task-orientation, persistence, and anxiety of mental hospital patients with high and low motivation. *Psychological Reports*, 1964, **14**, 18.

FISKE, D. W. Values, theory and the criterion problem. *Personnel Psychology*, 1951, **4**, 93–98.

FLEISHMAN, E. A. Leadership climate, human relations training, and supervisory behavior. *Personnel Psychology*, 1953, **6**, 205–222.

LEAVITT, H. J., & WHISLER, T. L. Management in the 1980's. *Harvard Business Review*, 1958, **36**, 41–48.

LOOMIS, J. L. Communication, the development of trust, and cooperative behavior. *Human Relations*, 1959, **12**, 305–315.

PORTER, L. W. Personnel management. *Annual Review of Psychology*, 1966, **17**, 395–422.

SAWYER, J. Measurement and prediction, clinical and statistical. *Psychological Bulletin*, 1966, **66**, 178–200.

SHARTLE, C. L. *Executive performance and leadership*. Englewood Cliffs, N. J.: Prentice-Hall, 1956.

VAUGHAN, J. A., & PORAT, A. M. Managerial reaction to computers. *Banking*, 1967, **59**(10), 119–122.

VON HAGEN, V. W. *Realm of the Incas*. New York: Mentor, 1957.

YONGE, K. A. The value of the interview: An orientation and a pilot study. *Journal of Applied Psychology*, 1956, **40**, 25–31.

(Received March 30, 1967)

journal of applied psychology
monograph supplement

Vol. 52, No. 1, Part 2 February 1968

EXPLORATIONS OF A THEORY OF VOCATIONAL CHOICE:

VI. A LONGITUDINAL STUDY USING A SAMPLE OF TYPICAL COLLEGE STUDENTS

JOHN L. HOLLAND [1]

American College Testing Program

Some hypotheses in Holland's theory were tested in a longitudinal study using large, diverse samples of college students. Using the appropriate operational definitions, students were classified as personality types and personality patterns and then examined to learn if they exhibited the characteristics suggested by the theory. This testing of concurrent and predictive relationships based on the theoretical formulations was generally positive. The testing of college influences upon vocational choice and satisfaction was accomplished by characterizing colleges by the Environmental Assessment Technique and students as personality types. These analyses of student-college interactions were partially successful. Taken together, the results support the usefulness of the theory.

This report is one of several studies undertaken to explore the usefulness of a theory of vocational choice and personality by testing some of its hypotheses (Holland, 1959, 1966b). The first and second monographs were longitudinal studies of bright students over 1–4 yr. periods (Holland, 1962, 1963). The third study was a longitudinal study of change in major field plans (Holland & Nichols, 1964). The fourth study examined some of the concurrent relationships between vocational preferences, occupational images, and self-ratings (Holland, 1963–64). The fifth report was another longitudinal study (Holland, 1964) that replicated some older hypotheses and tested some new and more remote hypotheses. It also fostered several revisions of the theory (Holland, 1959) and emphasized the need for a more explicit and comprehensive statement. Such a statement has since been published (Holland, 1966b).

Although these earlier studies support the theory, they were all based on samples of high-aptitude high school seniors or college students. The purpose of the present report was to test some of the hypotheses in the restatement of the theory and to do so in a large student sample approximating the typical college student. Consequently, this examination of the theory's hypotheses should be more explicit, and the findings should be applicable to a more general population.

[1] The author is indebted to the following persons for their counsel and skillful help in the planning and execution of this study: Thomas T. Frantz, Sandra W. Lutz, James M. Richards, Jr., William P. Alston, Theodore Cooper, Leonard P. Rand, and Lorraine M. Rand. He is also indebted to Alexander W. Astin, Leonard L. Baird, and Donald P. Hoyt for their constructive reading of an earlier draft of the manuscript.

© 1968 by the American Psychological Association, Inc.

TABLE 1

SAMPLES OF STUDENTS AND COLLEGES

College	Male	Female
Spring sample		
University of Alabama	186	201
Arkansas Polytechnic College	72	52
California State College at Hayward	62	74
Colorado State College	31	107
Southern Connecticut State College	89	273
Wesleyan University (Connecticut)	86	—
Indiana State University	126	164
Burlington Community College (Iowa)	79	31
Kansas State University	322	216
University of Kentucky	139	154
Westbrook Junior College (Maine)	—	103
William Jewell College (Missouri)	74	66
Plymouth State College (New Hampshire)	38	86
Glassboro State College (New Jersey)	121	393
University of North Dakota	89	123
Southeastern State College (Oklahoma)	97	61
Mount Mercy College (Pennsylvania)	—	104
Swarthmore College (Pennsylvania)	54	46
Black Hills State College (South Dakota)	60	38
University of Tennessee	205	184
Baylor University (Texas)	64	101
Fairmont State College (West Virginia)	93	85
Total	2087	2662
Fall sample		
California State College at Hayward	69	102
Chico State College (California)	109	173
Amherst College (Massachusetts)	238	—
University of Massachusetts	759	875
Baldwin-Wallace College (Ohio)	221	265
Cuyahoga Community College (Ohio)	180	156
Total	1576	1571

Briefly, the theory consists of several simple ideas and their more complex elaborations. First, we assume that we can characterize people by their resemblance to one or more personality types. The closer a person's resemblance to a particular type, the more likely it is that he will exhibit the personal traits and behaviors associated with that type. Second, we assume that the environments in which people live can be characterized by their resemblance to one or more model environments. Finally, we assume that the pairing of persons and environments leads to several outcomes that we can predict and understand from our knowledge of the personality types and the environmental models. These outcomes include vocational choice, vocational stability and achievement, personal stability, creative performance, and susceptibility to influence. These ideas, along with some operational definitions, are elaborated in *The Psychology of Vocational Choice* (Holland, 1966b). The following sections summarize the methods of collecting and analyzing the data and describe the findings.

METHODS OF ASSESSMENT

The data for the present study come from two American College Surveys described earlier by Richards, Holland, and Lutz (1966). Students were polled for their vocational aspirations and administered several scales and inventories. About 8 or 12 mo. later, students were polled again for their aspirations and satisfactions.

Students came from two college samples. The freshmen in the fall sample of six colleges were polled in the fall of 1964 and in May of 1965—an 8-mo. interval. The spring sample of college freshmen was polled in May of 1964 and again in May of 1965 when they were sophomores—a 12-mo. interval. The spring sample included 22 colleges. Both samples contained students with a great range of scholastic potential, vocational interests, and socioeconomic status. Table 1 shows the samples of colleges and students.

Because the follow-up data were obtained for only 39% of the spring sample and 44% of the fall sample, it was necessary to examine the differences between students with complete pre- and postdata and students with only initial data. Comparisons of students with complete and incomplete data in an earlier study using the same sample revealed that 18 out of 44 comparisons were significantly different for a variety of academic and nonacademic measures, but such mean differences *always* were less than one-quarter of a standard deviation (Richards et al., 1966). In the present study, comparisons of students with complete and incomplete data on the Vocational Preference Inventory scales, the most relevant variables in the present study, reveal that all mean differences are less than one-sixth of a standard deviation. The distributions of vocational choices are also similar. In short, there is no compelling evidence that the results would be substantially different if we had obtained complete data on the entire sample, although this occurrence is always possible. More important, attrition of the original sample is of little relevance because the main concern of the study is the testing of theoretical relationships in the *same* sample rather than the estimating of parameters for the original population.

A description of the assessment devices used in this study follows. The reliability estimates are based on the original spring samples of 6,289 male and 6,143 female college freshmen.

Vocational Preference Inventory (Sixth Revision). This personality and interest inventory is composed entirely of occupational titles (Holland, 1965). To take the inventory, a student indicates which occupations he likes and which he dislikes. Scores on only the following scales were used for this study: Realistic, Intellectual, Social, Conventional, Enterprising, and Artistic. Reliabilities (Kuder-Richardson 20) for these 14-item scales ranged from .83 to .89 for male college freshmen and from .76 to .89 for females.

Preconscious Activity Scale. This scale is an a priori scale developed to measure Kubie's (1958) notion of preconscious activity as a process in creative performance (Holland & Baird, in press). The Preconscious Activity Scale is a 38-item, true-false scale with reliabilities (K–R 20) of .72 and .68 for male and female college freshmen. The predictive

validities of this scale and its concurrent relationships with originality and interest measures imply that the Preconscious Activity Scale should be interpreted as an originality measure, especially in the fields of art, literature, and music.

Range of Competencies. Using a list of activities, students checked the activities which they claimed they could do well or competently. The assumption underlying these scales is that a large number of competencies is conducive to achievement generally and that competencies in a particular field are conducive to achievement in that field. Typical items from this list include: I have a working knowledge of *Roberts' Rules of Order*, I can dance, I am a good cook, I can make jewelry, I can read blueprints, I can read Greek, I can operate a tractor, I can use logarithm tables, etc. Competencies were categorized by three judges into several areas of competence: scientific, technical, business and clerical, social and educational, arts, and leadership and sales. Using unit weights, students were then scored for each kind of competency. The reliability for the competency scales used in the present study ranged from .48 to .87 and from .38 to .85 for men and women. The very low reliabilities for a few scales appear to result from the small number of items in such scales.

Interpersonal Competency Scale. This 20-item, a priori scale was modeled after the work of Foote and Cottrell (1955), who defined interpersonal competence as "acquired ability for effective interaction," and who outlined a program of research to study this concept. Scale items simply poll the subject for those factors which Foote and Cottrell believe to be conducive to, or typical of, interpersonal competency—good health, social experience and competencies, positive self-regard. The reliability (K–R 20) of the Interpersonal Competency Scale for male and female freshmen was .69 and .67, respectively.

Dogmatism Scale. This scale, developed by Rokeach (1956) to measure dogmatic and rigid thinking, consists of 40 true-false items dealing with beliefs and attitudes. (The first version by Rokeach is in multiple-choice form.) The reliability (K–R 20) was .77 for male freshmen and .75 for female freshmen.

Student Orientation Survey, Form C. Farber and Goodstein (1964) developed four a priori scales to assess the student orientations implied in Trow's (1960) student typology. These scales are Academic, Collegiate, Non-Conforming, and Vocational. The a priori scales were revised by an internal consistency item analysis to develop homogeneous, 10-item scales. Reliabilities (K–R 20) ranged from .39 to .45 for males and from .36 to .50 for females.

Other Descriptive Information. Students were polled for their educational and economic aspirations, their life goals, and their self-ratings. They were also asked to indicate their choice of vocation and field of training and to provide background information.

Students indicated the degree to which different life goals and achievements were "essential, very important, somewhat important, or of little importance" (e.g., being a religious person, making a contribution to scientific knowledge, being happy and content).

Using a list of traits and abilities such as originality, scholarship, and conservatism, students rated their personal traits and abilities on a 4-point scale—top 10%, above average, average, and below average.

PERSONALITY TYPES AND PATTERNS

The purpose of these analyses was to learn if a student characterized as a particular personality type, or as having a particular personality pattern, possesses the attributes hypothesized in the theory (Holland, 1966b). More concretely, if we determine a student's resemblance to each of the personality types—Realistic, Intellectual, Social, Conventional, Enterprising, and Artistic—by administering the Vocational Preference Inventory (VPI) and coding his profile of six scales according to the appropriate theoretical rules, will he exhibit the characteristics suggested by the theory? The following analyses show how well such student characteristics as personal traits, attitudes, aspirations, satisfactions, grades, and vocational roles can be predicted by using the theory to interpret the meaning of a student's VPI scale scores.

Student Traits and Orientations

In Tables 2–7, students have been classified by their highest VPI scale score, their two highest scores, and by their three highest scores.[2] Then, students with different high-point codes and profile patterns were compared for their competencies, life goals, self-ratings, attitudes, and personality traits.

Table 2, for men, reveals what students are like who peak on the various VPI scales. Generally, the student characteristics associated with individual VPI scales are expected ones. For example, Realistic peaks are associated with technical competencies and mechanical ability; Intellectual peaks are associated with scientific competencies, developing a scientific theory, mathematics ability, and an academic orientation. To facilitate study of these and succeeding tables, the "expected" high mean scale score for a given student attribute is shown in italics. And, al-

[2] The author is indebted to Sandra W. Lutz, Lorraine M. Rand, and Leonard P. Rand for reading the statement of the theory (Holland, 1966b) and establishing the predictions shown in these tables.

TABLE 2—*Continued*

Student characteristics		Highest VPI score						F
		Real (N= 885)	Int (N= 1774)	Soc (N= 845)	Conv (N= 428)	Ent (N= 798)	Art (N= 630)	
Nonconformist type	X	3.07	3.25	3.14	2.95	3.38	*3.74*	16.54
	SD	1.69	1.74	1.71	1.59	1.65	1.85	
Collegiate type	X	4.35	4.37	4.56	4.59	*5.05*	4.17	20.68
	SD	1.88	1.92	1.83	1.84	1.80	2.01	
Interpersonal competency	X	10.24	10.97	*11.78*	10.72	11.86	11.37	28.68
	SD	3.21	3.43	3.36	3.29	3.37	3.54	

Note.—All Fs are significant beyond the .05 level and the majority are significant beyond the .01 level. In Table 2, an F of 3.04 equals the .01 level.

The Ns for various VPI codes or scales in Tables 2–7 are larger than the same codes or scales in all other tables, because Tables 2–7 contain data for students with and without follow-up data. All other tables are based only on data for students with pre- and postdata.

Italicized means indicate theoretical predictions. If prediction is correct, italicized mean should be the highest X.

Abbreviations: Real or R = Realistic; Int or I = Intellectual; Soc or S = Social; Conv or C = Conventional; Ent or E = Enterprising; Art or A = Artistic.

though the highest mean is not always significantly different from other means for the same characteristic, 19 out of 25 predictions (76%) accord with the theoretical models.

Table 3 presents the student characteristics that go with a student's highest and second highest VPI scale scores. This table clearly indicates that students with the same high-point scale can still be distinguished by their second highest scale. In Table 3, for instance, men with a Realistic-Intellectual pattern are distinguished by their scientific competencies and goals of developing a useful product and scientific theory, whereas men with a Realistic-Social pattern are characterized by their social and educational competencies. In all, 69 out of 92 predictions (75%) are consistent with the theory.

In Table 4, students whose first two highest scales were the same were compared to see if they could still be distinguished by their third highest scale. To insure more reliable outcomes these analyses were performed only for the 20 three-variable patterns with the largest Ns. Despite this severe test of the validity of the personality patterns as assessed by the VPI, the F tests—although smaller in size and significant for fewer variables—reveal that the third scale still distinguishes one subgroup from another and usually in appropriate ways—25 out of 39 predictions are correct (64.1%).

The results for women are given in Tables 5, 6, and 7 for one-, two-, and three-variable VPI codes. In Table 5, 21 out of 25 predictions are correct (84%). In Table 6, 52 out of 69 predictions are correct (75.4%). And in Table 7, 24 out of 33 predictions are correct (72.7%).

The results presented in Tables 2–7 imply that the formulations for the personality types have some construct validity. Similarly, the interpretations of student profiles—VPI two-and-three-digit codes—by the application of the typological formulations possess similar validity.

Educational Aspiration, Satisfaction, and Achievement

In Tables 8, 9, and 10, the relations between personality types (high-point VPI codes) and satisfaction with college, educational aspiration, and college grades were examined. In Table 8, the chi-square tests show that high-point codes are not independent of a student's satisfaction with his college 8 mo. after he was assessed by the VPI. Students rated their satisfaction with their college by responding to the following item:

How satisfied are you with your final choice of a college?

This is the best possible college for me that I know of. 3

TABLE 3

RELATION OF TWO-DIGIT VPI CODES TO STUDENT CHARACTERISTICS (MEN)

Student characteristics		Two-digit VPI codes					F*
		RI (N=317)	RS (N=97)	RC (N=92)	RE (N=117)	RA (N=46)	
Competencies							
Scientific	X	*4.53*	3.22	3.54	3.39	3.46	10.43
	SD	2.40	2.10	2.24	2.22	2.22	
Technical	X	*15.09*	14.55	15.10	14.03	14.15	2.58
	SD	3.58	3.41	3.25	3.52	3.98	
Social and educational	X	4.29	*5.45*	4.38	4.68	5.26	4.02
	SD	2.81	2.69	2.64	2.83	3.04	
Artistic	X	5.29	4.71	4.93	5.68	*7.96*	4.46
	SD	4.35	3.71	4.37	5.51	5.75	
Life goals							
Developing useful product	X	*1.76*	1.39	1.65	1.44	1.39	7.28
	SD	.84	.67	.77	.68	.49	
Developing scientific theory	X	*1.76*	1.27	1.47	1.40	1.30	12.49
	SD	.81	.53	.74	.74	.51	
Becoming expert in finance and commerce	X	1.60	1.53	*1.79*	1.86	1.57	3.85
	SD	.76	.70	.89	.88	.71	
Active in religious affairs	X	2.53	*2.81*	2.62	2.74	2.89	2.85
	SD	.99	.98	.99	.96	.96	
Self-ratings							
Mechanical ability	X	*2.87*	2.58	2.82	2.57	2.52	5.38
	SD	.81	.88	.78	.79	.74	
Math ability	X	*2.49*	2.10	2.24	2.12	2.04	7.44
	SD	.93	.79	.79	.82	.93	
Personality and attitudinal scales							
Preconscious activity (originality)	X	15.50	14.93	14.42	14.07	*16.96*	4.73
	SD	4.60	4.11	4.24	4.34	4.92	
Collegiate type	X	4.13	4.94	4.09	*4.20*	4.74	4.67
	SD	1.85	1.85	1.76	1.85	1.69	
		IR (N=552)	IS (N=307)	IC (N=68)	IE (N=142)	IA (N=307)	
Competencies							
Technical	X	*14.47*	11.48	12.69	13.53	12.08	30.06
	SD	3.99	4.42	4.33	4.36	4.53	
Business	X	1.85	1.97	*2.31*	2.35	1.93	4.81
	SD	1.39	1.41	1.50	1.45	1.29	
Social and educational	X	4.81	*6.37*	5.69	6.37	5.44	17.69
	SD	2.87	2.86	2.99	2.96	3.07	
Artistic	X	5.95	7.04	6.13	8.30	*9.93*	27.57
	SD	5.29	4.97	6.24	6.35	5.86	
Leadership	X	3.71	4.66	4.21	*5.51*	4.61	12.93
	SD	2.80	3.07	2.97	3.27	3.13	
Life goals							
Developing useful product	X	*1.86*	1.61	1.63	1.69	1.74	5.37
	SD	.85	.79	.77	.79	.83	
Developing scientific theory	X	2.17	2.04	2.04	1.94	*2.24*	3.68
	SD	.94	.93	1.04	.83	.98	
Producing artwork	X	1.30	1.28	1.28	1.39	*1.62*	11.87
	SD	.64	.66	.56	.75	.87	

TABLE 3—*Continued*

Student characteristics		Two-digit VPI codes					F*
		IR (N=552)	IS (N=307)	IC (N=68)	IE (N=142)	IA (N=307)	
Becoming expert in finance and commerce	X	1.52	1.52	*2.09*	2.02	1.50	18.97
	SD	.74	.77	1.05	1.02	.79	
Active in religious affairs	X	2.55	*2.67*	2.71	2.47	2.27	7.08
	SD	1.00	1.01	1.04	1.02	1.06	
Executive responsible for work of others	X	2.21	2.23	2.68	*2.58*	2.11	12.01
	SD	.81	.86	.78	.83	.92	
Self-ratings							
Originality	X	2.33	2.42	2.41	2.49	*2.79*	22.33
	SD	.67	.68	.65	.68	.80	
Mechanical ability	X	*2.66*	2.04	2.31	2.25	2.28	27.44
	SD	.88	.86	.84	.87	.88	
Popularity	X	2.36	2.50	2.47	*2.60*	2.29	7.21
	SD	.65	.68	.76	.71	.75	
Understanding of others	X	2.58	*2.92*	2.68	2.79	2.77	12.36
	SD	.67	.71	.67	.67	.74	
Math ability	X	2.60	2.49	*2.88*	2.54	2.71	4.04
	SD	.91	.94	.88	.93	.94	
Personality and attitudinal scales							
Preconscious activity (originality)	X	16.94	18.10	14.99	16.26	*21.31*	63.34
	SD	4.22	4.72	4.60	4.15	4.68	
Dogmatism	X	17.51	17.51	*17.35*	18.32	16.50	2.75
	SD	5.69	6.02	5.97	6.58	5.55	
Vocational type	X	*5.02*	4.80	4.97	5.18	4.20	13.49
	SD	1.66	1.73	1.48	1.71	1.84	
Nonconformist type	X	3.14	3.08	3.24	3.46	*3.48*	3.22
	SD	1.70	1.65	1.84	1.73	1.75	
Collegiate type	X	4.28	4.63	4.40	*4.96*	3.90	10.01
	SD	1.86	1.83	1.93	1.97	1.92	
Interpersonal competency	X	10.48	*11.78*	11.44	11.82	10.66	10.51
	SD	3.16	3.35	2.90	3.39	3.81	
		SR (N=73)	SI (N=133)	SC (N=47)	SE (N=212)	SA (N=148)	
Competencies							
Scientific	X	4.18	*4.08*	2.83	3.55	3.05	5.86
	SD	2.47	2.22	1.94	2.40	2.36	
Technical	X	*13.05*	11.56	10.00	10.33	9.29	11.70
	SD	4.52	4.03	4.08	4.17	4.39	
Artistic	X	6.01	6.95	6.19	7.09	*9.77*	9.93
	SD	4.54	4.81	5.14	5.25	5.74	
Leadership	X	4.01	4.44	4.04	*5.16*	5.47	4.95
	SD	2.69	2.98	2.92	3.05	3.18	
Life goals							
Developing scientific theory	X	1.40	*1.53*	1.26	1.32	1.34	2.81
	SD	.61	.76	.52	.64	.63	
Producing artwork	X	1.37	1.32	1.26	1.25	*1.53*	3.97
	SD	.71	.68	.60	.60	.77	
Becoming expert in finance and commerce	X	1.55	1.49	*2.04*	1.71	1.43	7.11
	SD	.76	.77	1.01	.82	.67	
Executive responsible for work of others	X	2.07	2.26	2.30	*2.41*	2.22	2.38
	SD	.73	.89	.94	.87	.89	

TABLE 3—*Continued*

Student characteristics		Two-digit VPI codes					F*
		SR (N=73)	SI (N=133)	SC (N=47)	SE (N=212)	SA (N=148)	
Self-ratings							
Originality	X	2.16	2.37	2.02	2.41	*2.58*	7.24
	SD	.64	.77	.56	.72	.79	
Mechanical ability	X	*2.26*	2.15	1.64	1.77	1.68	11.60
	SD	.81	.88	.70	.85	.76	
Popularity	X	2.33	2.44	2.38	*2.64*	2.48	3.86
	SD	.70	.73	.76	.69	.64	
Math ability	X	2.00	*2.20*	2.40	2.00	1.92	3.78
	SD	.94	.95	.89	.82	.90	
Personality and attitudinal scales							
Preconscious activity (originality)	X	16.04	17.42	14.89	16.09	*19.91*	21.13
	SD	4.69	4.10	3.84	4.54	4.73	
Dogmatism	X	17.51	17.34	*19.53*	16.63	16.80	2.68
	SD	6.21	6.19	5.29	5.52	5.54	
Vocational type	X	*5.22*	4.62	5.23	5.06	4.24	6.91
	SD	1.85	1.78	1.74	1.74	1.77	
Collegiate type	X	4.67	4.31	4.53	*4.97*	4.13	5.57
	SD	1.84	1.71	1.61	1.83	1.82	
Interpersonal competency	X	11.04	11.44	10.79	*12.37*	12.19	4.62
	SD	3.27	3.19	3.34	3.29	3.22	
		CR (N=48)	CI (N=34)	CS (N=40)	CE (N=197)	CA (N=9)	
Competencies							
Scientific	X	3.37	*5.29*	2.87	3.21	3.11	7.20
	SD	1.78	2.84	2.17	2.16	2.33	
Technical	X	*13.62*	12.74	10.97	11.75	10.56	3.43
	SD	3.51	4.21	3.96	4.02	3.98	
Business	X	2.12	2.56	2.12	*2.55*	1.67	2.41
	SD	1.25	1.22	1.35	1.32	1.63	
Life goals							
Developing scientific theory	X	1.44	*1.76*	1.30	1.24	1.33	6.13
	SD	.61	.73	.56	.56	.67	
Becoming expert in finance and commerce	X	2.60	2.68	2.40	*3.06*	3.00	5.38
	SD	1.06	1.10	1.09	.89	1.05	
Self-ratings							
Originality	X	2.10	2.50	2.05	2.30	*2.67*	3.85
	SD	.47	.65	.63	.70	.67	
Mechanical ability	X	*2.44*	2.18	1.90	2.04	2.22	2.68
	SD	.91	.86	.89	.85	.79	
Math ability	X	2.46	*2.97*	2.50	2.40	2.44	2.94
	SD	.79	.92	1.00	.90	.83	
		ER (N=77)	EI (N=67)	ES (N=153)	EC (N=196)	EA (N=96)	
Competencies							
Scientific	X	3.53	*4.12*	3.59	3.21	3.10	2.51
	SD	2.33	2.03	2.67	2.29	2.18	
Technical	X	*13.96*	12.27	11.63	11.87	12.02	4.06
	SD	3.96	4.18	4.46	4.29	4.74	
Business	X	2.06	2.28	2.31	*2.61*	2.30	2.87
	SD	1.35	1.31	1.38	1.27	1.30	

TABLE 3—*Continued*

Student characteristics		Two-digit VPI codes					p*
		ER (N=77)	EI (N=67)	ES (N=153)	EC (N=196)	EA (N=96)	
Social and educational	X	5.45	5.91	*6.82*	5.80	6.31	6.14
	SD	2.82	3.27	2.80	2.64	3.04	
Artistic	X	6.05	6.30	7.93	5.81	*11.00*	15.74
	SD	5.29	4.39	6.11	5.26	6.40	
Leadership	X	4.30	5.10	*5.46*	5.17	5.80	2.73
	SD	3.12	3.27	3.25	2.95	3.22	
Life goals							
Producing artwork	X	1.27	1.33	1.32	1.29	*1.59*	3.55
	SD	.55	.76	.65	.72	.80	
Becoming expert in finance and commerce	X	2.03	2.43	2.02	*2.89*	2.26	20.75
	SD	.97	.95	1.00	.99	.99	
Self-ratings							
Originality	X	2.32	2.61	2.48	2.32	*2.72*	6.43
	SD	.81	.83	.66	.66	.77	
Mechanical ability	X	*2.43*	2.15	1.99	2.02	1.90	4.73
	SD	.86	.97	.90	.85	.85	
Popularity	X	2.47	2.63	*2.78*	2.61	2.73	3.05
	SD	.78	.75	.67	.70	.65	
Understanding of others	X	2.57	2.78	*2.90*	2.69	2.83	3.20
	SD	.76	.75	.74	.71	.72	
Math ability	X	2.13	*2.36*	1.92	2.20	1.86	5.25
	SD	.89	.93	.91	.90	.74	
Personality and attitudinal scales							
Preconscious activity (originality)	X	13.47	15.48	15.27	13.99	*18.14*	21.72
	SD	4.34	4.60	4.23	3.74	4.57	
Interpersonal competency	X	10.91	11.73	*12.37*	11.77	12.49	3.37
	SD	3.15	3.31	3.26	3.44	3.16	
		AR (N=41)	AI (N=145)	AS (N=161)	AC (N=6)	AE (N=146)	
Competencies							
Scientific	X	3.51	*4.42*	3.36	3.50	3.12	6.15
	SD	2.41	2.38	2.50	.96	2.30	
Technical	X	*12.24*	11.14	9.70	6.17	10.93	4.63
	SD	4.72	4.68	4.60	2.91	4.94	
Business	X	1.46	1.85	1.81	*2.33*	2.22	3.36
	SD	1.17	1.33	1.32	.94	1.50	
Social and educational	X	5.10	5.97	*6.77*	6.00	6.58	3.46
	SD	2.89	2.87	2.91	3.51	3.05	
Artistic	X	9.54	12.19	*13.08*	12.50	12.73	2.61
	SD	5.70	6.77	6.15	5.62	6.53	
Leadership	X	4.07	4.59	5.18	4.83	*5.72*	3.54
	SD	3.06	2.98	3.21	2.41	3.17	
Life goals							
Developing useful product	X	1.66	*1.52*	1.32	1.17	1.43	2.67
	SD	.84	.76	.65	.37	.70	
Developing scientific theory	X	1.37	*1.66*	1.33	1.50	1.38	4.91
	SD	.65	.85	.61	.76	.61	
Becoming expert in finance and commerce	X	1.66	1.46	1.45	*2.17*	2.02	11.55
	SD	.84	.74	.80	1.07	.97	
Executive responsible for work of others	X	2.00	2.00	2.03	3.00	*2.50*	8.65
	SD	.99	.89	.92	.58	.89	

TABLE 3—*Continued*

Student characteristics		Two-digit VPI codes					F*
		AR (N=41)	AI (N=145)	AS (N=161)	AC (N=6)	AE (N=146)	
Self-ratings							
Mechanical ability	X	*2.56*	2.05	1.74	1.50	1.88	8.47
	SD	.94	.87	.85	.50	.88	
Popularity	X	2.17	2.41	2.41	2.00	*2.54*	2.48
	SD	.79	.71	.78	.58	.78	
Understanding of others	X	2.73	2.93	*3.06*	2.83	2.82	2.67
	SD	.83	.73	.75	.69	.76	
Math ability	X	2.17	*2.30*	1.81	2.67	1.99	6.32
	SD	.99	1.05	.87	.47	.88	
Personality and attitudinal scales							
Preconscious activity (originality)	X	21.15	*23.30*	22.80	19.50	20.77	6.13
	SD	5.11	5.20	5.36	3.69	4.39	
Vocational type	X	*4.51*	3.86	4.04	5.00	4.53	3.13
	SD	1.84	1.82	2.00	1.53	1.85	
Collegiate type	X	3.56	4.00	4.04	4.50	*4.77*	4.86
	SD	2.02	1.84	2.07	2.43	1.90	
Interpersonal competency	X	10.05	11.29	*11.43*	9.83	12.09	3.46
	SD	3.62	3.62	3.07	1.77	3.43	

Note.—See note of Table 2.
* All F ratios are significant beyond .05 level.

This is a good college for me, but there are a few others that I think are better. 2
This is only a fair college, and there are many others which would probably suit me better. 1

The results for men in Table 8 conform closely with the theoretical statement (Holland, 1966b, p. 48): satisfaction should be highest for the Intellectual type, and in descending order, Social, Artistic, Conventional, Enterprising, and Realistic. The rank-order correlation between the actual mean satisfaction scores and the mean expected scores equals .83 ($p < .05$, one-tailed test). At the extremes, the male Intellectual type is most satisfied with his college and the Realistic type is least satisfied. The results for women follow an unpredictable pattern and do not conform to expectation.

In Table 9, student educational aspirations are not independent of high-point codes for each sex ($p < .001$). The two rank-order correlations (each rho = .77) between the observed and predicted medians for the six types are positive, but they are not significant. The expected level of educational aspiration by personality type is identical with the theoretical ordering reported earlier for satisfaction with college and also constitutes the expected order for level of academic achievement in Table 10.

In Table 10, college grades, for men, are not independent of personality types ($p < .05$). For women, grades and types are independent of one another ($p > .05$). Although grades, for men, are not independent of types, the rho of .71 between the observed and expected median grade for each type is not significant. (A rho of .83 is needed for a one-tailed test at the 5% level.)

Vocational Role Preferences

In a follow-up survey, about 8 or 12 mo. after the administration of the VPI, students responded to the following item:

What special role would you like to play within your first occupational choice?

Being a practitioner of my occupation.
Training or teaching others about my occupation.
Leading or supervising people.
Doing research in my field.
Acting as a consultant or expert to others.
Undecided.
Other role.

TABLE 4

RELATION OF THREE-DIGIT VPI CODES TO STUDENT CHARACTERISTICS (MEN)

Student characteristics		Three-digit VPI codes				F*
		RIS ($N=59$)	RIC ($N=59$)	RIE ($N=59$)	RIA ($N=44$)	
Competencies						
Social and educational	*X*	*5.31*	3.17	4.93	4.09	7.42
	SD	2.65	2.62	2.83	2.40	
Artistic	*X*	5.14	3.88	6.24	*7.82*	7.57
	SD	3.88	3.95	4.44	5.02	
Leadership	*X*	3.54	2.39	*3.86*	3.41	3.65
	SD	2.26	2.42	2.61	2.90	
Life goals						
Producing artwork	*X*	1.22	1.25	1.47	*1.55*	2.88
	SD	.52	.54	.81	.84	
Becoming expert in finance and commerce	*X*	1.61	*1.58*	1.88	1.41	3.74
	SD	.76	.74	.78	.58	
Self-ratings						
Popularity	*X*	2.24	2.10	*2.44*	2.30	3.16
	SD	.65	.60	.59	.55	
Personality and attitudinal scales						
Preconscious activity (originality)	*X*	14.73	15.22	17.05	*17.80*	5.38
	SD	4.33	4.69	4.66	4.35	
Collegiate type	*X*	3.88	3.95	*4.85*	4.41	3.53
	SD	1.70	1.61	1.94	2.04	
Interpersonal competency	*X*	*10.64*	9.14	10.73	10.00	3.54
	SD	3.00	2.97	3.02	2.90	
		IRS ($N=115$)	IRC ($N=76$)	IRE ($N=89$)	IRA ($N=86$)	
Competencies						
Social and educational	*X*	*5.56*	4.04	5.29	4.58	5.30
	SD	2.85	2.62	3.02	2.71	
Artistic	*X*	6.30	4.99	5.76	*7.56*	3.55
	SD	5.43	4.61	4.89	5.59	
Leadership	*X*	3.90	3.14	*4.33*	3.45	3.03
	SD	2.80	2.49	2.87	2.58	
Life goals						
Developing scientific theory	*X*	1.98	2.30	2.12	*2.30*	2.97
	SD	.85	.89	.83	.97	
Becoming expert in finance and commerce	*X*	1.49	*1.66*	1.80	1.30	7.73
	SD	.65	.77	.84	.61	
Executive responsible for work of others	*X*	2.13	2.08	*2.47*	2.03	5.96
	SD	.77	.74	.72	.80	
Self-ratings						
Originality	*X*	2.27	2.29	2.25	*2.56*	4.40
	SD	.62	.68	.60	.69	
Math ability	*X*	2.47	*2.96*	2.54	2.52	5.29
	SD	.88	.82	.96	.89	
Personality and attitudinal scales						
Preconscious activity (originality)	*X*	16.44	16.32	15.88	*19.09*	11.06
	SD	4.28	4.14	3.92	3.87	
Vocational type	*X*	4.91	*5.25*	5.45	4.78	2.97
	SD	1.77	1.62	1.69	1.57	
Collegiate type	*X*	4.40	3.79	*4.61*	4.19	3.00
	SD	1.91	1.74	1.78	1.81	
Interpersonal competency	*X*	*11.25*	9.75	10.69	10.58	3.59
	SD	2.92	3.05	3.01	3.40	

TABLE 4—*Continued*

Student characteristics		Three-digit VPI codes				F[a]
		IAR (N=93)	IAS (N=94)	IAC (N=7)	IAE (N=49)	
Competencies						
Technical	X	*12.87*	10.89	11.57	12.90	3.91
	SD	4.35	4.34	3.33	4.48	
Business	X	1.74	1.88	*3.14*	2.22	3.83
	SD	1.29	1.14	.83	1.37	
Social and educational	X	4.55	*6.37*	4.71	6.37	7.31
	SD	3.04	3.01	1.67	2.76	
Leadership	X	3.76	5.01	4.43	*6.10*	6.45
	SD	2.75	3.24	2.92	3.39	
Life goals						
Developing useful products	X	*1.87*	1.47	1.71	1.92	5.15
	SD	.87	.65	.70	.92	
Becoming expert in finance and commerce	X	1.38	1.30	*2.57*	1.90	12.16
	SD	.65	.71	.90	.95	
Executive responsible for work of others	X	1.87	2.11	2.86	*2.49*	6.28
	SD	.89	.94	.64	.97	
Self-ratings						
Originality	X	2.83	*2.90*	2.00	2.82	2.83
	SD	.84	.73	.53	.82	
Mechanical ability	X	*2.54*	2.07	2.57	2.10	5.21
	SD	.92	.84	1.05	.84	
Popularity	X	2.14	2.32	3.00	*2.55*	5.21
	SD	.74	.66	1.69	.70	
Personality and attitudinal scales						
Vocational type	X	*4.20*	3.70	4.71	4.73	3.98
	SD	1.89	1.80	1.48	1.50	
Interpersonal competency	X	9.76	*11.00*	9.43	12.37	5.65
	SD	3.68	4.05	1.59	3.24	
		ISR (N=84)	ISC (N=12)	ISE (N=60)	ISA (N=60)	
Competencies						
Business	X	1.70	1.67	*2.45*	1.88	3.74
	SD	1.21	1.43	1.50	1.42	
Artistic	X	6.30	8.08	7.17	*9.25*	4.47
	SD	4.40	5.47	4.51	5.45	
Life goals						
Producing artwork	X	1.18	1.42	1.12	*1.53*	5.34
	SD	.47	.86	.45	.88	
Self-ratings						
Originality	X	2.40	2.50	2.28	*2.70*	4.30
	SD	.64	.65	.61	.71	
Mechanical ability	X	*2.20*	2.42	1.85	1.98	2.95
	SD	.94	.86	.73	.76	
Personality and attitudinal scales						
Preconscious activity (originality)	X	17.61	17.92	16.32	*21.28*	12.49
	SD	4.35	3.86	4.93	4.77	

Note.—See note of Table 2.
[a] All F ratios are significant beyond .05 level.

TABLE 5

RELATION OF ONE-DIGIT VPI CODES TO STUDENT CHARACTERISTICS (WOMEN)

Student characteristics		One-digit VPI codes						F[a]
		Real (N = 31)	Int (N = 554)	Soc (N = 3033)	Conv (N = 271)	Ent (N = 148)	Art (N = 1284)	
Competencies								
Scientific	X	3.52	*4.51*	2.83	2.28	2.47	2.87	65.84
	SD	2.49	2.42	2.14	1.81	2.14	2.21	
Technical	X	*9.45*	5.63	4.78	4.90	5.26	4.97	19.72
	SD	4.10	3.31	3.07	3.13	3.69	3.23	
Business	X	2.00	1.94	2.03	*2.48*	2.32	2.02	9.29
	SD	1.11	1.24	1.25	1.20	1.39	1.24	
Social and educational	X	6.19	6.67	*8.01*	6.62	7.61	7.56	40.19
	SD	2.72	2.90	2.38	2.50	2.73	2.63	
Arts	X	8.23	9.68	10.07	7.76	10.87	*13.89*	106.34
	SD	5.37	5.34	5.50	4.58	6.11	6.41	
Leadership	X	3.55	4.24	5.07	4.16	*5.53*	4.97	12.75
	SD	2.91	2.97	3.07	3.07	3.32	3.11	
Life goals								
Developing useful product	X	1.29	*1.35*	1.19	1.20	1.23	1.24	9.05
	SD	.45	.63	.50	.57	.55	.58	
Developing scientific theory	X	1.42	*1.74*	1.19	1.16	1.21	1.23	85.15
	SD	.55	.88	.52	.49	.55	.60	
Producing artwork	X	1.48	1.56	1.45	1.41	1.66	*1.97*	65.66
	SD	.84	.87	.79	.75	.94	1.06	
Becoming expert in finance and commerce	X	1.45	1.34	1.37	*1.85*	1.72	1.33	31.58
	SD	.66	.66	.70	1.01	.91	.68	
Active in religious affairs	X	2.52	2.74	*3.03*	2.96	2.62	2.72	26.08
	SD	.95	1.03	.94	.93	.99	1.05	
Executive responsible for work of others	X	1.84	1.84	2.01	2.14	*2.32*	1.85	15.84
	SD	.77	.84	.88	.91	.90	.87	
Self-ratings								
Originality	X	2.13	2.31	2.24	2.12	2.45	*2.59*	54.22
	SD	.71	.71	.66	.67	.68	.74	
Mechanical ability	X	*2.39*	1.85	1.57	1.67	1.68	1.63	21.09
	SD	.90	.82	.70	.77	.78	.75	
Popularity	X	2.26	2.23	2.36	2.27	*2.40*	2.33	4.55
	SD	.67	.65	.63	.66	.69	.65	
Understanding of others	X	2.58	2.71	*2.84*	2.60	2.78	2.84	9.74
	SD	.71	.73	.67	.63	.67	.70	
Math ability	X	2.10	*2.36*	1.90	2.32	1.95	1.86	38.58
	SD	.82	.99	.85	.97	.83	.88	
Conservatism	X	2.03	2.22	2.18	*2.31*	2.20	2.10	6.68
	SD	.54	.71	.63	.63	.74	.69	
Personality and attitudinal scales								
Preconscious activity (originality)	X	17.52	19.81	18.12	14.49	17.07	*22.20*	215.60
	SD	4.76	4.61	4.49	4.30	4.94	4.50	
Dogmatism	X	16.68	16.26	17.05	*18.26*	17.68	16.60	6.56
	SD	6.50	5.52	5.53	5.30	5.09	5.48	
Academic type	X	4.29	*5.00*	4.57	4.27	4.39	4.87	10.78
	SD	1.99	2.03	1.92	1.82	1.93	1.95	
Vocational type	X	*4.71*	4.16	4.57	5.01	4.58	4.01	30.24
	SD	1.51	1.65	1.67	1.79	1.67	1.68	
Nonconformist type	X	3.23	2.82	2.74	2.61	2.94	*3.11*	11.10
	SD	1.75	1.55	1.58	1.53	1.88	1.74	
Collegiate type	X	4.42	4.56	5.22	4.87	*5.18*	4.69	19.98
	SD	1.76	1.99	1.97	1.95	1.97	2.06	
Interpersonal competency	X	10.06	10.63	*11.78*	10.51	11.51	11.62	18.93
	SD	3.39	3.30	3.20	3.35	3.45	3.26	

Note.—See note of Table 2.
[a] All F ratios are significant beyond .05 level.

TABLE 6

RELATION OF TWO-DIGIT VPI CODES TO STUDENT CHARACTERISTICS (WOMEN)

Student characteristics		Two-digit VPI codes					F*
		IR (N=32)	IS (N=234)	IC (N=25)	IE (N=14)	IA (N=150)	
Competencies							
Technical	X	*7.09*	5.40	6.96	4.57	5.29	3.82
	SD	3.42	3.07	3.39	2.87	3.25	
Business	X	1.78	2.02	*2.52*	2.57	1.65	5.03
	SD	.99	1.22	1.20	1.24	1.19	
Social and educational	X	4.62	*7.09*	6.72	8.21	6.27	7.34
	SD	2.81	2.67	2.88	2.51	3.02	
Artistic	X	7.44	9.58	7.24	10.43	*10.57*	3.84
	SD	4.41	5.57	3.95	4.84	5.31	
Leadership	X	3.12	4.45	4.72	*6.29*	3.89	3.83
	SD	2.52	3.03	2.89	2.99	2.83	
Life goals							
Producing artwork	X	1.47	1.42	1.40	2.21	*1.71*	5.37
	SD	.83	.78	.57	.94	.90	
Becoming expert in finance and commerce	X	1.22	1.35	*1.60*	1.93	1.27	4.57
	SD	.54	.67	.80	.80	.60	
Executive responsible for work of others	X	1.50	1.87	2.20	*2.43*	1.75	4.90
	SD	.71	.80	.89	1.05	.83	
Self-ratings							
Originality	X	2.16	2.26	2.04	2.50	*2.43*	3.02
	SD	.75	.65	.77	.73	.73	
Mechanical ability	X	*2.62*	1.75	2.08	2.07	1.81	9.29
	SD	.89	.80	.80	.88	.77	
Popularity	X	2.09	2.29	2.16	*2.71*	2.14	3.79
	SD	.52	.66	.54	.70	.63	
Understanding of others	X	2.34	*2.80*	2.48	3.29	2.71	5.76
	SD	.69	.67	.75	.70	.77	
Math ability	X	2.72	2.25	*2.84*	2.43	2.35	3.46
	SD	.84	.94	.92	1.05	1.01	
Personality and attitudinal scales							
Preconscious activity (originality)	X	18.50	19.21	16.84	18.93	*21.69*	11.91
	SD	4.15	4.56	4.44	4.33	3.95	
Academic type	X	4.44	5.10	3.92	5.07	*5.13*	2.77
	SD	1.85	1.98	1.49	1.22	2.14	
Collegiate type	X	3.66	4.69	4.72	*6.14*	4.43	4.51
	SD	2.07	1.93	1.95	1.88	1.92	
Interpersonal competency	X	9.78	*11.16*	10.72	12.36	10.08	4.17
	SD	3.67	3.14	3.58	2.84	3.13	
		SR (N= 21)	SI (N= 444)	SC (N= 360)	SE (N= 536)	SA (N= 1077)	
Competencies							
Scientific	X	3.10	*4.00*	2.31	2.54	2.73	43.97
	SD	2.31	2.34	1.95	2.06	2.01	
Technical	X	*6.95*	5.25	5.09	4.63	4.50	8.78
	SD	3.37	3.43	3.11	2.94	2.92	
Business	X	2.05	1.95	*2.22*	2.12	1.95	4.43
	SD	1.09	1.23	1.23	1.29	1.24	

TABLE 6—*Continued*

Student characteristics		Two-digit VPI codes					F*
		SR (N= 21)	SI (N= 444)	SC (N= 360)	SE (N= 536)	SA (N= 1077)	
Social and educational	X	7.29	7.91	7.74	*8.34*	8.10	4.55
	SD	2.00	2.49	2.24	2.42	2.36	
Artistic	X	8.29	9.27	8.26	9.80	*11.52*	32.62
	SD	5.48	5.29	4.66	5.24	5.68	
Leadership	X	4.90	4.95	4.71	*5.40*	5.18	3.19
	SD	3.10	2.97	3.06	3.09	3.06	
Life goals							
Developing useful product	X	1.19	*1.24*	1.15	1.16	1.19	2.38
	SD	.66	.54	.43	.44	.52	
Developing scientific theory	X	1.19	*1.36*	1.17	1.15	1.16	12.85
	SD	.66	.68	.50	.46	.49	
Producing artwork	X	1.33	1.43	1.33	1.41	*1.58*	9.01
	SD	.84	.77	.65	.75	.86	
Becoming expert in finance and commerce	X	1.29	1.32	*1.56*	1.44	1.28	14.06
	SD	.63	.67	.80	.74	.61	
Active in religious affairs	X	2.76	3.03	*3.14*	2.95	3.06	2.82
	SD	.97	.94	.89	.97	.94	
Executive responsible for work of others	X	1.76	2.05	2.12	*2.12*	1.94	5.78
	SD	.81	.92	.80	.88	.86	
Self-ratings							
Originality	X	2.00	2.22	2.08	2.23	*2.32*	9.76
	SD	.53	.65	.56	.63	.71	
Mechanical ability	X	*2.00*	1.69	1.60	1.54	1.52	6.72
	SD	.87	.76	.70	.71	.68	
Popularity	X	2.29	2.34	2.23	*2.43*	2.37	5.86
	SD	.55	.65	.51	.65	.65	
Understanding of others	X	2.57	2.83	2.76	*2.87*	2.86	2.73
	SD	.90	.69	.65	.64	.68	
Math ability	X	1.86	*2.03*	1.97	1.82	1.84	5.82
	SD	.89	.87	.84	.80	.85	
Conservatism	X	2.10	2.23	*2.23*	2.10	2.20	3.68
	SD	.53	.65	.61	.58	.66	
Personality and attitudinal scales							
Preconscious activity (originality)	X	15.71	18.52	15.62	16.91	*19.75*	86.67
	SD	4.59	4.19	3.73	4.28	4.24	
Dogmatism	X	16.14	16.95	*18.21*	17.21	16.91	4.30
	SD	5.44	5.37	5.30	5.62	5.45	
Academic type	X	3.57	*4.87*	4.46	4.35	4.68	7.14
	SD	2.01	1.82	1.83	1.93	1.95	
Vocational type	X	*5.14*	4.37	5.05	4.68	4.40	13.58
	SD	1.32	1.69	1.61	1.66	1.60	
Nonconformist type	X	2.76	2.74	2.52	2.87	*2.73*	2.56
	SD	1.80	1.54	1.51	1.67	1.59	
Collegiate type	X	5.62	5.09	5.25	*5.66*	5.03	10.14
	SD	1.65	1.96	1.76	1.94	2.02	
Interpersonal competency	X	11.24	11.66	11.41	*12.22*	11.86	4.20
	SD	3.32	3.30	3.01	3.10	3.19	
		CR (N=3)	CI (N=15)	CS (N=117)	CE (N=53)	CA (N=21)	
Competencies							
Scientific	X	4.00	*3.33*	2.04	2.34	2.19	2.46
	SD	2.16	1.96	1.76	1.83	1.71	

TABLE 6—*Continued*

Student characteristics		Two-digit VPI codes					F
		CR ($N=3$)	CI ($N=15$)	CS ($N=117$)	CE ($N=53$)	CA ($N=21$)	
Technical	X	*11.67*	4.27	4.63	4.55	5.29	4.54
	SD	2.62	3.07	3.01	2.69	2.69	
Life goals							
Becoming expert in finance and commerce	X	1.33	1.73	1.74	*2.26*	1.86	2.85
	SD	.47	.85	.92	1.18	.89	
Active in religious affairs	X	3.00	2.27	*3.00*	3.00	2.71	2.44
	SD	.82	.57	.92	.97	.98	
Executive responsible for work of others	X	1.33	1.73	2.04	*2.53*	2.29	4.44
	SD	.47	.77	.91	.88	.76	
Self-ratings							
Mechanical ability	X	*2.67*	1.80	1.50	1.74	1.95	3.83
	SD	.47	.65	.65	.83	.90	
Understanding of others	X	1.67	2.33	*2.68*	2.62	2.71	2.85
	SD	.47	.47	.60	.68	.70	
Personality and attitudinal scales							
Nonconformist type	X	5.67	2.40	2.60	2.81	*2.67*	3.10
	SD	1.89	.88	1.49	1.69	1.61	
		ER ($N=2$)	EI ($N=5$)	ES ($N=44$)	EC ($N=13$)	EA ($N=25$)	
Competencies							
Artistic	X	5.50	7.00	10.98	9.69	*15.24*	3.89
	SD	4.50	3.63	6.52	5.33	5.52	
Leadership	X	2.50	5.60	*6.32*	3.38	7.12	3.59
	SD	.50	1.62	3.54	2.87	2.86	
Life goals							
Developing scientific theory	X	1.00	*2.20*	1.32	1.23	1.16	2.96
	SD	0.00	.98	.63	.42	.61	
Personality and attitudinal scales							
Preconscious activity (originality)	X	13.00	17.40	15.95	14.92	*20.68*	5.67
	SD	2.00	5.82	4.41	5.01	3.97	
		AR ($N=13$)	AI ($N=175$)	AS ($N=773$)	AC ($N=13$)	AE ($N=118$)	
Competencies							
Scientific	X	3.00	*4.00*	2.72	3.69	2.12	17.30
	SD	2.00	2.55	2.06	2.84	1.87	
Technical	X	*7.77*	5.45	4.79	3.62	5.09	4.71
	SD	3.74	3.62	3.05	2.59	3.37	
Social and educational	X	6.00	7.25	*7.86*	6.15	7.18	5.63
	SD	2.72	2.61	2.56	2.90	2.78	
Artistic	X	9.85	14.79	*14.14*	11.77	13.82	2.40
	SD	5.92	7.21	6.26	7.36	6.28	
Leadership	X	3.46	4.69	5.06	3.23	*5.61*	3.52
	SD	2.98	3.24	3.06	2.61	3.08	
Life goals							
Developing scientific theory	X	1.31	*1.41*	1.20	1.15	1.19	4.83
	SD	.61	.73	.56	.36	.56	
Producing artwork	X	1.85	2.24	*1.92*	2.00	2.02	3.33
	SD	1.03	1.13	1.04	.78	1.02	
Becoming expert in finance and commerce	X	1.08	1.27	1.30	*1.69*	1.52	4.43
	SD	.27	.63	.64	.91	.85	

TABLE 6—*Continued*

Student characteristics		Two-digit VPI codes					F*
		AR (N=13)	AI (N=175)	AS (N=773)	AC (N=13)	AE (N=118)	
Active in religious affairs	X	2.23	2.49	*2.82*	2.69	2.76	4.69
	SD	.97	1.07	1.02	.82	1.05	
Executive responsible for work of others	X	1.69	1.70	1.84	1.92	*2.09*	3.79
	SD	.72	.87	.86	1.00	.91	
Self-ratings							
Mechanical ability	X	*2.23*	1.83	1.56	1.46	1.54	7.52
	SD	.70	.81	.72	.50	.71	
Understanding of others	X	2.46	2.81	*2.89*	2.46	2.69	4.15
	SD	.63	.72	.71	.50	.66	
Personality and attitudinal scales							
Preconscious activity (originality)	X	23.08	*23.64*	22.39	18.69	21.23	8.31
	SD	5.23	4.37	4.24	4.14	4.43	
Collegiate type	X	4.08	4.47	4.76	3.38	*4.85*	2.48
	SD	2.34	2.01	2.09	2.13	1.96	
Interpersonal competency	X	10.31	11.25	*11.88*	9.08	11.57	4.21
	SD	3.00	3.32	3.27	3.36	3.04	

Note.—See note of Table 2.
* All F ratios are significant beyond .05 level.

The analysis in Table 11 was performed to learn if role preferences for students with different VPI codes conform to theoretical expectations (Holland, 1966b, p. 65, Table 7). The distributions of student role preferences for each VPI code were rank ordered and correlated with the expected ranks. The results in Table 11 are all positive, but only rhos of .90 or 1.00 are significant at the 5% level for a one-tailed test. In short, students with some high-point codes have very predictable choices of vocational role, notably, Realistic, Social, and Conventional men, and Social and Conventional women.

To summarize, the evidence in Tables 2–11 suggests that the personality types are useful models for interpreting and organizing a broad range of information about students. Despite some undesirable overlapping of characteristics across types, the majority of expected typological characteristics conformed to the theoretical formulations. However, some revisions of the theory are needed for the role expectations for each type, for the assignment of a few traits to types, and for the application of the theory to women.

PREDICTION OF VOCATIONAL CHOICE AND OTHER CRITERIA

The theory assumes that Realistic types select Realistic occupations and that Intellectual types select Intellectual occupations, etc. The following analyses test this main hypothesis with different assessment devices—the VPI and the student's initial expressed vocational choice. In addition, several closely related hypotheses were examined: the importance of a student's preference of vocational role in prediction, the prediction of *degrees* of change in a student's vocational aspiration, and the improvement of predictive efficiency by using various combinations of predictors or special attributes of the predictors.

Classification Scheme

For the following analyses and all other analyses in this report, student vocational choices were categorized according to the classification scheme developed earlier: six categories for men—Realistic, Intellectual, Social, Conventional, Enterprising, and Artistic—and eight categories for women—Intellectual, Social-Intellectual, Social-Conven-

TABLE 7

RELATION OF THREE-DIGIT VPI CODES TO STUDENT CHARACTERISTICS (WOMEN)

Student characteristics		Three-digit VPI codes				F*
		SIR (N=22)	SIC (N=34)	SIE (N=78)	SIA (N=175)	
Competencies						
Technical	X	*7.45*	6.12	5.41	5.05	3.51
	SD	5.29	3.54	3.60	3.16	
Artistic	X	8.91	7.97	9.62	*10.58*	2.63
	SD	4.07	5.99	5.56	5.38	
Personality and attitudinal scales						
Preconscious activity (originality)	X	16.73	16.74	18.18	*19.55*	7.23
	SD	4.79	4.13	3.94	4.02	
Vocational type	X	*4.36*	5.06	4.72	4.11	4.11
	SD	1.55	1.76	1.66	1.75	
Interpersonal competency	X	12.86	10.41	*12.32*	11.71	3.77
	SD	3.49	2.86	2.84	3.27	
		SER (N=15)	SEI (N=43)	SEC (N=130)	SEA (N=173)	
Competencies						
Scientific	X	1.93	*3.35*	2.09	2.48	4.95
	SD	1.57	1.96	1.68	2.09	
Technical	X	*6.93*	4.77	4.39	4.49	3.67
	SD	3.26	2.55	2.88	2.87	
Social and educational	X	7.93	8.81	8.06	*8.89*	3.53
	SD	2.72	2.47	2.52	2.17	
Artistic	X	7.60	9.88	8.54	*11.17*	7.75
	SD	3.67	4.04	5.05	5.36	
Life goals						
Developing useful product	X	1.67	*1.21*	1.09	1.17	7.36
	SD	.87	.55	.31	.46	
Developing scientific theory	X	1.47	*1.23*	1.12	1.12	3.50
	SD	.81	.56	.39	.40	
Self-ratings						
Math ability	X	1.60	*1.93*	1.92	1.69	3.15
	SD	.80	.79	.81	.70	
Personality and attitudinal scales						
Preconscious activity (originality)	X	15.40	17.09	15.62	*18.20*	10.99
	SD	3.26	4.51	3.98	3.96	
Academic type	X	5.00	*5.14*	4.27	4.38	2.70
	SD	2.07	1.80	2.02	1.86	
Interpersonal competency	X	11.13	12.53	11.85	*12.80*	3.29
	SD	2.75	2.76	3.07	3.10	
		SAR (N=16)	SAI (N=206)	SAC (N=93)	SAE (N=443)	
Competencies						
Scientific	X	1.81	*3.46*	2.76	2.62	9.71
	SD	1.24	2.14	2.29	1.89	
Social and educational	X	7.25	7.80	7.98	*8.58*	7.23
	SD	2.63	2.45	2.29	2.19	

TABLE 7—*Continued*

Student characteristics		Three-digit VPI codes				F[a]
		SAR (N=16)	SAI (N=206)	SAC (N=93)	SAE (N=443)	
Artistic	X	9.12	*11.43*	11.16	12.52	3.81
	SD	3.94	6.16	6.18	5.56	
Leadership	X	3.87	5.13	4.94	*5.77*	4.72
	SD	3.33	3.02	3.32	2.97	
Self-ratings						
Originality	X	2.31	*2.38*	2.15	2.42	3.79
	SD	.58	.72	.67	.69	
Math ability	X	2.00	*1.90*	2.04	1.73	4.68
	SD	1.00	.83	.89	.84	
Personality and attitudinal scales						
Preconscious activity (originality)	X	19.87	*20.81*	18.26	19.83	8.51
	SD	3.44	4.11	3.71	4.12	
Collegiate type	X	4.31	4.76	4.80	*5.42*	7.23
	SD	1.96	1.96	2.02	2.00	
Interpersonal competency	X	12.25	11.59	11.27	*12.36*	4.95
	SD	3.27	3.25	3.07	3.00	
		ASR (N=22)	ASI (N=211)	ASC (N=34)	ASE (N=339)	F[a]
Competencies						
Scientific	X	2.59	*3.12*	2.47	2.42	5.59
	SD	1.87	2.07	1.61	1.98	
Business	X	2.09	1.86	*2.79*	2.08	6.19
	SD	1.24	1.14	1.11	1.22	
Social and educational	X	7.23	7.72	8.00	*8.23*	2.63
	SD	2.13	2.47	2.49	2.49	
Leadership	X	3.91	4.80	4.56	*5.60*	5.04
	SD	2.92	2.77	3.42	3.13	
Life goals						
Executive responsible for work of others	X	1.55	1.73	2.03	*1.91*	3.42
	SD	.72	.81	.71	.89	
Self-ratings						
Popularity	X	2.05	2.28	2.41	*2.43*	4.26
	SD	.56	.62	.65	.66	
Math ability	X	1.77	*2.01*	1.91	1.73	4.72
	SD	.85	.88	.95	.82	
Personality and attitudinal scales						
Preconscious activity (originality)	X	24.00	*23.05*	20.12	22.21	6.21
	SD	4.01	3.97	3.98	4.39	
Collegiate type	X	4.32	4.59	4.82	*5.08*	3.08
	SD	1.79	2.14	2.01	1.98	

Note.—See note of Table 2.
[a] All F ratios are significant beyond .05 level.

TABLE 8

RELATION OF SATISFACTION WITH COLLEGE AND PERSONALITY TYPE (VPI HIGH-POINT CODES) FOR FALL SAMPLE

VPI high-point code	Men (N = 1353)[a]			Women (N = 1383)[b]		
	Low	Middle	High	Low	Middle	High
Real	19.3 (22)	46.5 (53)	33.3 (38)	0.0 (0)	0.0 (0)	100.0 (3)
Int	8.2 (45)	50.5 (277)	41.2 (226)	12.3 (28)	49.8 (113)	37.4 (85)
Soc	9.4 (20)	50.9 (108)	38.7 (82)	12.4 (93)	44.9 (338)	42.6 (321)
Conv	12.4 (13)	43.8 (46)	43.8 (46)	4.0 (2)	48.0 (24)	48.0 (24)
Ent	15.0 (29)	53.9 (104)	31.1 (60)	13.9 (5)	61.1 (22)	25.0 (9)
Art	10.2 (19)	49.5 (92)	39.2 (73)	17.4 (55)	47.6 (151)	34.7 (110)

Note.—Chi-square calculated from frequencies shown in parentheses. Abbreviations: Real = Realistic; Int = Intellectual; Soc = Social; Conv = Conventional; Ent = Enterprising; Art = Artistic.
[a] $\chi^2 = 21.97$; $df = 10$; $p < .02$.
[b] $\chi^2 = 21.27$; $df = 10$; $p < .02$.

TABLE 9

RELATION OF EDUCATIONAL ASPIRATION TO PERSONALITY TYPES (HIGH-POINT VPI CODES)

VPI high-point code	Men (N=1328)[a]		Women (N=1363)[b]	
	BA or less (%)	Advanced training (%)	BA or less (%)	Advanced training (%)
Real	43.8	56.3	100.0	0.0
Int	17.6	82.4	48.2	51.8
Soc	29.1	70.9	58.6	41.4
Conv	47.6	52.4	80.0	20.0
Ent	31.6	68.4	68.6	31.4
Art	16.9	83.1	45.9	55.1

Note.—See note of Table 8 for abbreviations.
[a] $\chi^2 = 75.26$; $df = 5$; $p < .001$.
[b] $\chi^2 = 38.74$; $df = 5$; $p < .001$.

TABLE 10

RELATION OF COLLEGE GRADES TO PERSONALITY TYPES (VPI HIGH-POINT CODES)

VPI high-point code	Men (N=1359)[a]		Women (N=1386)[b]	
	≤C	≥C+	≤C	≥C+
Realistic	50.9	49.1	33.3	66.7
Intellectual	43.4	56.6	42.3	57.7
Social	50.0	50.0	45.3	54.7
Conventional	55.2	44.8	52.0	48.0
Enterprising	53.9	46.1	55.6	44.4
Artistic	44.1	55.9	44.5	55.5

[a] $\chi^2 = 11.41$; $df = 5$; $p < .05$.
[b] $\chi^2 = 3.68$; *ns*.

tional, Social-Enterprising, Social-Artistic, Conventional, Enterprising, and Artistic (Holland, 1966a). Tables 12 and 13 indicate the assignment of vocational choices to vocational classes for men and women. Students selected their career choices from a coded list of 99 careers. All classifications, as well as the scoring and establishing the highest VPI scale codes, were performed by a computer. We assumed, then, that the classifications were performed with perfect reliability.

TABLE 11

CORRELATIONS (RANK ORDER) BETWEEN PREDICTED AND OBSERVED ROLE PREFERENCES FOR THE PERSONALITY TYPES

	Personality types (VPI high-point codes)					
	Real	Int	Soc	Conv	Ent	Art
Fall sample (1359 men; 1386 women)						
Men	.90*	.60	.90*	1.00**	.50	.20
Women	.75	.80	.90*	.90*	.50	.60
Spring sample (1773 men; 2336 women)						
Men	.70	.60	.90*	1.00**	.50	.18
Women	.50	.60	.90*	1.00**	.68	.10

* p <.05, one-tailed test.
** p <.01, one-tailed test.

TABLE 12

PSYCHOLOGICAL CLASSIFICATION SCHEME FOR VOCATIONS (MEN)

Realistic class		
Agricultural Science	Forestry	Industrial Engineering
Architecture	Geography	Mechanical Engineering
Civil Engineering	Industrial Arts Education	Trade & Industrial Education
Farming		
Intellectual class		
Aeronautical Engineering	Engineering; General, Other	Oceanography
Anthropology	Engineering Science	Other Biological Science Fields
Astronomy, Astrophysics	Geology, Geophysics	Other Health Fields
Biochemistry	Mathematics Education	Pharmacy
Biology	Mathematics, Statistics	Physical Therapy
Botany	Medical Technology	Physics
Chemical Engineering	Medicine	Physiology
Chemistry	Metallurgical Engineering	Veterinary Science
Dentistry	Military Service	Zoology
Electrical Engineering	Natural Science Education	
Social class		
Clinical Psychology	Experimental & General Psychology	Industrial & Personnel Psychology
Counseling & Guidance	Foreign Language Education	Physical Education, Recreation, & Health
Education, General, & Other Specialties	Foreign Service	Social Work
Education of Exceptional Children	General Social Sciences	Sociology
Educational Psychology	History	Theology, Religion
Elementary Education	History Education	
Conventional class		
Accounting	Business Education	Finance
Enterprising class		
Economics	Other Business & Commerce	Public Relations
Law	Political Science	Purchasing
Management	Public Administration	Sales
Marketing		
Artistic class		
Art	General Humanities	Music Education
Art Education	Journalism, Radio-TV, Communication	Other Fine & Applied Arts
Drama	Literature	Philosophy
English, Creative Writing	Music	Speech
English Education		

Vocational Choice and VPI [3]

The prediction of a student's final vocational choice from his first choice 8 mo. earlier is shown in Table 14 for men in the fall sample. The percentage of correct predictions varies from 63.2 to 71.2%. The total number of correct predictions always exceeds base-rate expectations and cannot be attributed to chance.

[3] These analyses were reported earlier as ACT Research Report No. 18 (Holland & Lutz, 1967).

TABLE 13

PSYCHOLOGICAL CLASSIFICATION SCHEME FOR VOCATIONS (WOMEN)

Intellectual class		
Agricultural Science Architecture Biochemistry Biology	Chemistry Mathematics, Statistics Medicine Natural Science Education	Other Biological Sciences Physics Veterinary Medicine Zoology
Social-intellectual class		
Clinical Psychology Dentistry Experimental & General Psychology Mathematics Education	Medical Technology Nursing Other Health Fields Pharmacy	Physical Therapy Political Science, Government, International Relations Theology, Religion
Social-conventional class		
Business Education	Clerical, Office Work	Secretarial Science
Social-enterprising class		
Educational Psychology Management, Business Administration	Purchasing	Sales
Social-artistic class		
Counseling & Guidance Education, General & Other Specialties Education of Exceptional Children Elementary Education English Education History	History Education Home Economics Home Economics Education Housewife Law Modern Foreign Language Education	Physical Education, Recreation, & Health Public Relations, Advertising Social Science Social Work, Group Work Sociology Speech
Conventional class		
Accounting		
Enterprising class		
Marketing		
Artistic class		
Art Art Education Drama English, Creative Writing Foreign Service	Journalism, Radio-TV, Communication Library Science, Archival Science Literature	Modern Foreign Language Music Music Education Other Fine & Applied Arts Philosophy

TABLE 14

PREDICTION OF FINAL VOCATIONAL CHOICE FROM FIRST VOCATIONAL CHOICE
(FALL SAMPLE, MEN, N = 1359)

First vocational choice	Final vocational choice									
	Real	Int	Soc	Conv	Ent	Art	Und	% hits	NR	N
Realistic (Real)	*106*	13	4	1	12		14	67.9	6	156
Intellectual (Int)	31	*355*	21	14	29	8	57	67.6	10	525
Social (Soc)	3	8	*110*	1	12	3	19	68.8	4	160
Conventional (Conv)		1	2	*42*	7		7	71.2		59
Enterprising (Ent	4	5	15	10	*155*	4	32	67.1	6	231
Artistic (Art)	1	2	8		6	*48*	10	63.2	1	76
Undecided (Und)	4	14	13	4	18	4	53		1	111
No response (NR)		10	8	5	8		9		1	41

Note.—The percentages of hits in Table 14 and succeeding tables are based on the total N in each row and are a conservative estimate, because students who are "undecided" or give "no response" in the follow-up survey are included in the total N.

The prediction of a student's final vocational choice from his highest score (high-point code) among six VPI scores, obtained 8 mo. earlier, is given in Table 15. The percentage of correct predictions ranges only from 21.5 to 51.4%. In this instance, it is clear that simply asking the student is superior to using the VPI, although both methods of assessing personality types have predictive validity.

The discrepancy between the efficiency of a student's expressed choice and his VPI scores becomes even greater when we select a subsample of students whose first two vocational choices, at the time of initial testing, fall in the same vocational class: for example, physics and chemistry, education and social work, art and literature, etc. The exact questionnaire items for this analysis were as follows:

My present career choice is: (Select the appropriate number from the list of coded careers and curricula)

If I could not have my first choice (above) I would select the following occupation: (Select the appropriate number from this list)

TABLE 15

PREDICTION OF FINAL VOCATIONAL CHOICE FROM VPI HIGH-POINT CODE
(FALL SAMPLE, MEN, N = 1359)

VPI high-point code	Final vocational choice									
	Real	Int	Soc	Conv	Ent	Art	Und[a]	% hits	NR[b]	N
Realistic (Real)	*43*	28	9	7	8		17	37.7	2	114
Intellectual (Int)	70	*282*	45	9	43	14	70	51.4	16	549
Social (Soc)	8	32	*78*	9	34	11	37	36.8	3	212
Conventional (Conv)	11	19	4	*33*	23		13	31.4	2	105
Enterprising (Ent)	9	16	17	18	*98*	2	29	50.8	4	193
Artistic (Art)	8	31	28	1	41	*40*	35	21.5	2	186

Note.—To make single predictions from the VPI, it was necessary to omit students whose two highest scores were tied. This occurrence then necessitated the omission of students with tied profiles from the tables of expressed choice so that the comparisons of the VPI and expressed choice are based on identical samples. If, however, "expressed" choice predictions are based on all students (with *and* without VPI ties), the total predictive efficiency shown in Tables 14 and 17 vary only by 1% or less.

[a] Und = undecided.
[b] NR = no response.

TABLE 16

PREDICTION OF FINAL VOCATIONAL CHOICE FOR STUDENTS WHOSE FIRST TWO CHOICES FALL IN THE SAME CLASS
(FALL SAMPLE, MEN, N = 586)

First and second vocational choices	Final vocational choice									
	Real	Int	Soc	Conv	Ent	Art	Und	% hits	NR	N
Realistic (Real)	*36*	2	1		6		1	76.6	1	47
Intellectual (Int)	18	*239*	5	4	10	4	21	78.1	5	306
Social (Soc)			*47*		3	1	9	78.3		60
Conventional (Conv)				*12*	2			85.7		14
Enterprising (Ent)	1		2	2	*69*	2	9	78.4	3	88
Artistic (Art)			3		3	*22*	1	73.3	1	30
Undecided (Und)		4	1		2	1	5			13
No response (NR)		9	3	3	6		7			28

Table 16 reveals that subgrouping students whose first two choices belong to the same vocational class results in a substantial gain in predictive efficiency. Correct predictions range from 73.3 to 85.7% as contrasted with 63.2 to 71.2 for the *total* sample of men. The subgroup of students whose first two choices fall in different vocational classes are, as expected, less predictable from their first choice than students whose first two choices belong to the same class. Although the percentages of hits are lower—53.0 to 66.7—these percentages still exceed the percentages obtained by the use of the VPI—21.5 to 51.4.

The results for women in the fall sample are presented in Tables 17, 18, and 19. The main findings for women parallel those for men. The analyses for women are identical to those for men except for the use of a special classification scheme developed for women in which the social class is divided into four subclasses. This subdivision was made to distribute women more equally among the classes or types. Expressed vocational choices predict 34.8–83.8% of later choices (Table 17).

The VPI in Table 18 predicts final choices, on the average, less efficiently than asking the student (0.0–81.9%). In Table 19, the formation of a subsample of women whose first

TABLE 17

PREDICTION OF FINAL VOCATIONAL CHOICE FROM FIRST VOCATIONAL CHOICE
(FALL SAMPLE, WOMEN, N = 1386)

First vocational choice	Final vocational choice											
	Int	Soc Int	Soc Conv	Soc Ent	Soc Art	Conv	Ent	Art	Und	% hits	NR	Total
Intellectual (Int)	*54*	19	2	2	18	2		4	15	42.5	11	127
Social Intellectual (Soc Int)	10	*183*	3	6	42	1	1	10	7	66.3	13	276
Social Conventional (Soc Conv)			*17*		1	1			3	73.9	1	23
Social Enterprising (Soc Ent)			2	*8*	10	1	1			34.8	1	23
Social Artistic (Soc Art)	4	20	4	4	*553*	1		25	37	83.8	12	660
Conventional (Conv)	2		1		1	*5*				55.6		9
Artistic (Art)	1	3	2		34			*73*	19	52.1	8	140
Undecided (Und)	1	7	1	1	28			11	24		12	85
No response (NR)	5	4			16			2	7		9	43

TABLE 18

PREDICTION OF FINAL VOCATIONAL CHOICE FROM VPI HIGH-POINT CODE
(FALL SAMPLE, WOMEN, N = 1386)

VPI high-point code	Final vocational choice											
	Int	Soc Int	Soc Conv	Soc Ent	Soc Art	Conv	Ent	Art	Und[a]	% hits	NR[b]	Total
Realistic (Real)		2	1							0.0		3
Intellectual (Int)	*46*	64	4	2	55	1		17	25	20.3	13	227
Social (Soc)	14	*111*	*18*	*7*	*481*	3	1	38	54	81.9	26	753
Conventional (Conv)	5	9	5	2	12	*5*		3	5	10.0	4	50
Enterprising (Ent)		7		3	16	1		2	5	0.0	2	36
Artistic (Art)	12	43	4	7	139	1	1	*65*	23	20.5	22	317

[a] Und = undecided.
[b] NR = no response.

TABLE 19

PREDICTION OF FINAL VOCATIONAL CHOICE FOR STUDENTS WHOSE FIRST TWO CHOICES FALL IN THE SAME CLASS
(FALL SAMPLE, WOMEN, N = 545)

First and second vocational choice	Final vocational choice										
	Int	Soc Int	Soc Conv	Soc Ent	Soc Art	Conv	Art	Und	% hits	NR	Total
Intellectual (Int)	*25*	7			1		1	3	65.8	1	38
Social Intellectual (Soc Int)	2	*85*	2	2	5	1		1	85.0	2	100
Social Conventional (Soc Conv)			*8*			1		1	72.7	1	11
Social Enterprising (Soc Ent)			1	*1*					50.0		2
Social Artistic (Soc Art)	2	8	1		*299*		8	13	88.5	7	338
Conventional (Conv)	1										1
Artistic (Art)		1	1		8		*23*	2	60.5	3	38
Undecided (Und)								1			1
No response (NR)	1	3			5		1	5		1	16

two choices belong to the same class yields the most efficient predictions—50.0 to 88.5%. Again Table 19 demonstrates that students whose first two choices belong to the *same* vocational class are more predictable than students whose choices belong to different classes (32.6–78.9%). (The data for the latter analysis are not presented.)

Because of the controversial character of the results obtained for the fall sample, the same analyses were performed for the spring sample with similar results. These analyses involve a longer interval of time (1 yr. instead of 8 mo.) and a different stage of col-

TABLE 20

SUMMARY FOR SPRING SAMPLE

Kinds of prediction	% correct predictions	
	Men (N = 1773)	Women (N = 2336)
Expressed vocational choice—total	68.7	78.2
VPI—highest scale	45.1	59.6
Expressed vocational choice—same	82.5	86.4
Expressed vocational choice—different	64.2	71.9

lege life (end of freshman year to end of sophomore year, as opposed to beginning of freshman year to end of freshman year). Table 20 is a summary of the main results. Without exception, the results in Table 20 replicate what we found earlier—various categorizations of expressed vocational choice are clearly and substantially superior to the VPI.

TABLE 21

COMPARISON OF THE PREDICTIVE VALUES OF HIGH-POINT VPI CODE AND PREFERRED ROLE, AND INITIAL VOCATIONAL CHOICE AND PREFERRED ROLE (FALL SAMPLE)

	Preferred vocational role											
	Und., Oth., NR*		Practitioner		Teacher		Leader		Researcher		Consultant	
	f	% hits	*f*	% hits	*f*	% hits	*f*	% hits	*f*	% hits	*f*	% hits
Men (N = 1359)												
VPI												
Real	45	22.2	34	*47.1*	6	50.0	11	54.5	9	33.3	9	55.6
Int	155	36.1	154	57.8	35	45.7	23	30.4	160	*66.3*	22	36.4
Soc	73	27.4	77	39.0	27	*48.1*	13	46.2	8	37.5	14	42.9
Conv	42	21.4	32	34.4	3	66.7	11	9.1	4	25.0	13	*69.2*
Ent	63	47.6	69	53.6	5	0.0	30	*63.3*	5	60.0	21	42.9
Art	71	16.9	65	29.2	9	33.3	20	15.0	13	7.7	8	25.0
Total		*30.5*		*46.9*		*43.5*		*38.9*		*58.8*		*44.8*
Vocational choice												
Real	39	61.5	60	*71.7*	7	57.1	16	75.0	21	57.1	13	84.6
Int	129	57.4	171	71.3	37	64.9	22	40.9	149	*76.5*	17	70.6
Soc	36	52.8	57	77.2	24	*87.5*	19	57.9	13	46.2	11	81.8
Conv	20	50.0	17	82.4	1	100.0	2	0.0	2	100.0	17	*88.2*
Ent	66	59.1	87	72.4	4	75.0	43	*76.7*	7	42.9	24	58.3
Art	22	50.0	34	64.7	11	72.7	4	100.0	3	66.7	2	50.0
Total		*56.7*		*71.5*		*72.6*		*65.1*		*71.3*		*73.8*
Women (N = 1386)												
VPI												
Real	1	0.0	2	*0.0*								
Int	65	15.4	76	19.7	17	0.0	12	8.3	51	*33.3*	6	50.0
Soc	205	74.1	347	87.0	80	*76.3*	57	93.0	21	66.7	43	81.4
Conv	17	23.5	18	5.6	4	0.0	5	0.0	3	0.0	3	*0.0*
Ent	10	0.0	13	0.0	2	0.0	7	*0.0*	3	0.0	1	0.0
Art	88	22.7	143	19.6	37	27.0	18	16.7	11	0.0	20	20.0
Total		*48.2*		*57.8*		*50.7*		*57.6*		*34.8*		*57.5*
Vocational choice												
Int	34	20.6	38	63.2	5	20.0	1	100.0	44	40.9	5	60.0
Soc Int	62	61.3	129	74.4	26	46.2	29	55.2	24	75.0	6	50.0
Soc Conv	4	50.0	13	76.9	4	100.0					2	50.0
Soc Ent	1	0.0	345	1.7	3	0.0	4	75.0			4	25.0
Soc Art	141	75.9			75	77.3	57	86.0	16	62.5	37	86.5
Conv	7	57.1	1	100.0			1	0.0				
Ent												
Art	39	35.9	59	62.7	21	52.4	5	80.0	2	0.0	14	50.0
Total		*59.7*		*29.7*		*64.2*		*75.3*		*53.5*		*69.1*

Note.—The Ns in this table represent all available data. The Ns for totals and subtotals differ from the total sample Ns and are not internally consistent because students who had tied VPI profiles, gave unclassifiable choices, or failed to respond have been omitted. The italicized percentages indicate the "expected" vocational role for each class or type.

* Undecided, Other, No Response.

Vocational Choice, VPI, and Role

The following analyses (see Footnote 3) were performed to see how well we could predict a student's later vocational choice when his preferred vocational role was considered along with his first vocational choice or his VPI scores. Because special vocational roles are associated with each environment (Holland, 1966b, pp. 64–65), and by implication with each type, we assumed that more efficient predictions would be obtained for students whose personality type and preference for vocational role were congruent. For example, an Intellectual type who prefers the role of researcher should be more predictable than the same type who prefers the role of supervisor.

Table 21, a summary of 24 tables, presents the predictive efficiencies of a student's highest VPI score *and* his role preference, and his expressed vocational choice *and* his role preference. For men, students who prefer the roles of practitioner, teacher, leader, researcher, or consultant are more likely to give the same or closely related vocational choice 8 mo. later than are students who fail to respond, or who respond "undecided" or "other role." Or about two-thirds of the comparisons reveal that male students with explicit role preferences are more stable than students with ambiguous preferences. The results for women in Table 21 imply that vocational role adds little or nothing to a woman's initial expression of vocational choice.

The italicized percentages within the six classes indicate the "expected" vocational role for each class or type (Holland, 1966b, p. 65). The efficiency of these theoretically derived predictions is unclear because many cells in Table 21 are too small to yield stable percentages. If the "total" percentages in Table 21 are, however, used as a base, the "expected" percentages of hits, for men, exceed the corresponding base percentages 10 out of 10 times! For instance, students with peaks on Social and preferences for the role of teacher have a hit rate of 48.1% whereas the hit rate for "teacher" generally is 43.5. The results for women are again ambiguous for several reasons. Both the VPI and the classification scheme create many small cells, and it is unclear how to derive a single set of explicit predictions from the vocational-choice classification for women.

TABLE 22

RELATION OF PERSONALITY TYPES (HIGH-POINT VPI CODE) TO STABILITY OF VOCATIONAL CHOICE (FALL SAMPLE)

First and final vocational choice	High-point VPI code						
	Real	Int	Soc	Conv	Ent	Art	Total
Men (N = 1359)[a]							
No change	50.9	50.6	42.0	49.5	40.9	47.3	644
Intraclass	14.0	16.9	16.0	13.3	21.8	14.5	226
Interclass	35.1	32.4	42.0	37.1	37.3	38.2	489
Women (N = 1386)[b]							
No change	33.3	48.0	51.0	58.0	52.8	47.6	735
Intraclass	.0	6.6	13.7	6.0	13.9	13.9	185
Interclass	67.7	45.4	27.3	36.0	33.3	38.5	463

[a] $\chi^2 = 14.17$; $df = 10$; $p < .20$.
[b] $\chi^2 = 40.67$; $df = 10$; $p < .001$.

The efficiency of the predictions obtained for expressed choice *and* vocational role approximates those obtained by subgrouping students whose first two choices belong to the same vocational class.

In short, we can predict vocational choices most accurately by one of two simple methods: (1) ask the student about his first two vocational choices, or (2) ask him once about his vocational intentions and then ask him for his preferred vocational role. Either of these methods is almost twice as efficient as the VPI.

Degrees of Change

In Table 22, the relation of personality type (high-point VPI code) to degrees of change in vocational choice is examined. In this table, "no change" equals giving identical responses over an 8-mo. interval, "intraclass change" equals giving two occupational choices that belong to the same occupational class (see Tables 12 and 13), and "interclass change" equals giving choices that belong to different classes.

Table 22 reveals that, for women, some personality types are more stable than others—notably the Social and Conventional types. The results for men are not significant and

fail to support earlier findings for high-aptitude students (Holland, 1962).

Personality Patterns

The theory assumes that the consistency and the homogeneity of a student's VPI pattern, or personality pattern, are indicative of the stability of his initial vocational choice (Holland, 1966b, p. 43f, p. 76f); that is, consistent patterns should be more stable than inconsistent patterns, and homogeneous profiles should be more stable than heterogeneous profiles.

Tests of the consistency hypothesis are shown in Tables 23 and 24 for men in the fall and spring samples. In 11 out of 12 comparisons, the percentages of hits for students with the same VPI code favor the students with consistent codes. Consistent Realistic types are more stable than inconsistent Realistic types, etc. However, only 4 out of the 12 comparisons are significant at the 5% level (one-tailed test). A new test with a larger N is needed to clarify these results. The results for women are not only not significant but not worth reexamination in the context of this theory.

In Table 25, students were categorized as having VPI profiles of high, middle, or low homogeneity (the difference between the highest and lowest VPI raw score), then compared for their stability of vocational choice. For men, the results clearly support the hypothesis —homogeneity is positively related to stability, or, spiked profiles are stable and flat profiles are unstable ($p < .05$). The results for women are not significant.

Another test of the homogeneity hypothesis consisted of examining the stability of vocational choice for students whose VPI profiles had (1) "no ties"—a single scale was higher than any other scale, (2) the two highest scales were "tied"—both had equal scores, or (3) the three or more highest scales were tied. The results of this analysis, shown in Table 26, were statistically significant and according to expectation for men. Again, the results for women are not significant.

TABLE 23

CONSISTENCY OF THE PERSONALITY PATTERN (TWO-DIGIT VPI CODE) AND STABILITY OF VOCATIONAL CHOICE AFTER 1 YEAR (FALL SAMPLE, MEN, $N = 1359$)

VPI high-point code	Final vocational choice									
	Real	Int	Soc	Conv	Ent	Art	Und[a]	% hits	NR[b]	Total
Consistent two-digit VPI codes										
Realistic (Real)	*34*	22	4	7	4		10	40.9	2	83
Intellectual (Int)	55	*182*	12	3	22	9	38	54.3[c]	14	335
Social (Soc)	3	13	*49*	9	22	9	23	37.7	2	130
Conventional (Conv)	8	11	4	*30*	22		9	34.9	2	86
Enterprising (Ent)	6	9	15	16	*76*	2	26	49.4	4	154
Artistic (Art)	6	28	27		35	*38*	32	22.6	2	168
Inconsistent two-digit VPI codes										
Realistic	9	6	5		4		7	29.0		31
Intellectual	15	*100*	33	6	21	5	32	46.7[c]	2	214
Social	5	19	*29*		12	2	14	35.4	1	82
Conventional	3	8		*3*	1		4	15.8		19
Enterprising	3	7	2	2	*22*		3	56.4		39
Artistic	2	3	1	1	6	*2*	3	11.1		18

[a] Und = undecided.
[b] NR = no response.
[c] $\chi^2 = 3.00$; $df = 1$; $p < .05$, one-tailed test.

TABLE 24

CONSISTENCY OF THE PERSONALITY PATTERN (TWO-DIGIT VPI CODE) AND STABILITY OF VOCATIONAL CHOICE AFTER 1 YEAR (SPRING SAMPLE, MEN, $N = 1773$)

VPI high-point code	Final vocational choice									
	Real	Int	Soc	Conv	Ent	Art	Und	% hits	NR	Total
Consistent two-digit codes										
Realistic (Real)	*71*	57	15	11	21	1	15	36.8[a]	2	193
Intellectual (Int)	62	*245*	33	6	25	19	23	58.2[b]	8	421
Social (Soc)	6	18	*84*	2	24	14	13	51.5	2	163
Conventional (Conv)	8	5	6	*34*	35	1	2	36.6	2	93
Enterprising (Ent)	11	15	24	15	*78*	8	8	48.8[c]	1	160
Artistic (Art)	5	28	36	1	28	*59*	18	33.0	4	179
Inconsistent two-digit VPI codes										
Realistic	*26*	16	25	2	12	1	18	25.5[a]	2	102
Intellectual	13	*115*	38	7	29	7	14	51.1[b]	2	225
Social	15	16	*67*	1	22	5	10	47.5	5	141
Conventional	4	6		*4*	4	1	2	19.0		21
Enterprising	14	12	8	3	*15*	2	4	25.9[c]		58
Artistic	1	8	2	1	2	*2*	1	11.8		17

[a] $\chi^2 = 3.82$; $df = 1$; $p < .05$, one-tailed test.
[b] $\chi^2 = 2.99$; $df = 1$; $p < .05$, one-tailed test.
[c] $\chi^2 = 9.05$; $df = 1$; $p < .01$, one-tailed test.

Although this second analysis did not use the standard definition of homogeneity—the difference between highest and lowest VPI scale scores—it does suggest that the relative sharpness of a profile, as assessed by the crude process of counting the number of tied high points in a profile, is indicative of a student's stability of vocational choice. Taken together, the analyses in Tables 25 and 26 imply that the construct, homogeneity, has some validity for men, but not for women.

TABLE 25

HOMOGENEITY OF THE PERSONALITY PATTERN (VPI PROFILE) AND STABILITY OF VOCATIONAL CHOICE (FALL SAMPLE)

Homogeneity	Men ($N = 1359$)[a]				Women ($N = 1386$)[b]			
	Hits		Misses		Hits		Misses	
	f	%	*f*	%	*f*	%	*f*	%
High	201	46.4	232	53.6	248	52.9	221	47.1
Middle	205	42.6	276	57.4	276	55.7	220	44.3
Low	168	37.8	277	62.2	209	49.6	212	50.4

[a] $\chi^2 = 6.81$; $df = 2$; $p < .05$.
[b] $\chi^2 = 3.31$; *ns*.

Combinations of Consistency, Homogeneity, and Role

It follows from the theory that students whose VPI profiles are consistent and homogeneous, and who express a vocational role preference other than "undecided, no response, or other role" will be more predictable than students who have inconsistent and heterogeneous profiles, and who fail to express a role preference. To learn if the effects

TABLE 26

HOMOGENEITY OF A STUDENT'S VPI PROFILE AND STABILITY OF HIS VOCATIONAL CHOICE (FALL SAMPLE)

Homogeneity	Men ($N = 1402$)[a]				Women ($N = 1427$)[b]			
	Hits		Misses		Hits		Misses	
	f	%	*f*	%	*f*	%	*f*	%
No ties	816	67.6	391	32.4	893	71.0	365	29.0
2 ties	77	64.7	42	35.3	80	66.7	40	33.3
3 or more	40	52.6	36	47.4	34	69.4	15	30.6
Total	933		469		1007		420	

[a] $\chi^2 = 7.43$; $df = 2$; $p < .05$.
[b] $\chi^2 = 1.04$; *ns*.

TABLE 27

PREDICTION OF FINAL VOCATIONAL CHOICE FROM STUDENT'S VPI PROFILE AND ROLE PREFERENCE

VPI profile						Samples (males only)					
						Fall		Spring		Fall-Int[c]	
Consistency		Homogeneity		Preferred role		% hits	*f*	% hits	*f*	% hits	*f*
C[a]	+	Hi	+	Explicit	=	51.8	305	53.4	509	68.2	107
I[b]	+	Hi	+	Explicit	=	52.2	115	53.0	185	54.4	49
C	+	Lo	+	Explicit	=	44.2	330	45.0	420	51.2	63
I	+	Lo	+	Explicit	=	45.0	160	37.3	233	47.5	56
C	+	Hi	+	Ambiguous	=	33.3	126	39.5	124	46.4	13
I	+	Hi	+	Ambiguous	=	26.2	42	32.1	56	30.0	10
C	+	Lo	+	Ambiguous	=	31.8	195	39.1	156	27.1	13
I	+	Lo	+	Ambiguous	=	25.6	86	28.9	90	19.4	6

[a] Consistency.
[b] Inconsistency.
[c] Fall–Intellectual.

of consistency, homogeneity, and vocational role are cumulative so that their occurrence in combination results in more efficient predictions, all possible combinations of these variables were tried. Table 27 presents the results of these analyses for men in the fall and spring samples.

The results suggest that consistency, homogeneity, and vocational role preference cumulate to improve the prediction of vocational choice, but their relative contributions appear very unequal. Explicit role preference appears to contribute most to the prediction, followed next by homogeneity, and last by consistency.

The predictions in Table 27 can be improved by controlling for VPI high-point code, that is, by recalculating the analyses in Table 27 for a single high-point code at a time. These analyses were done, but they require large *N*s to insure reliable percentages. For the largest subsample—male students who peaked on the Intellectual scale of the VPI—the results are especially clear and are shown in the last column of Table 27. Similar analyses for the smaller remaining subsamples produced similar but less-consistent results. (It is instructive to note that the best predictive combinations for men and women in Table 27 are either equal to, or less than, the predictive validity of a student's first vocational choice in Tables 14 and 17.) Aside from their theoretical value, the results in Table 27 reinforce some common interpretations about vocational counseling clients; namely, the row at the top of the table characterizes the people vocational counselors rarely see, whereas the bottom row characterizes the kinds of people counseling agencies usually deal with; that is, clients who produce confused interest profiles and who are unclear or uncertain about the vocational role they want to play.

To summarize, the predictive data in Tables 14 through 27 are strong evidence that the types and their associated constructs—vocational role preferences, consistency, and homogeneity—generally yield efficient and theoretically relevant predictions for men but only occasionally for women. At this point, the efficiency of self-expression of vocational choice suggests that a more thorough exploration of the general usefulness of this technique for defining types is desirable. Similarly, a redefinition of consistency could be achieved by the use of a person's first and second vocational choice, and this new definition appears now to have more potential than the definition of consistency by the VPI.

INSTITUTIONAL INFLUENCES

The following analyses were performed to learn if a college's influence upon a student's choice of vocation could be predicted from

the theory by characterizing students and colleges in terms of personality types and environmental models, and then testing the appropriate hypotheses (Holland, 1966b, p. 72f). More explicitly, the analyses are concerned with the stability of vocational choice, or whether or not a student switches from one vocational class to another. A second group of analyses was performed to learn if a student's satisfaction with his college is predictable within the context of the theory.

Assessment of Types and Environments

In these analyses, a student's initial vocational choice was used as an index of his personality type (see Tables 12 and 13) and as a control for student input. The student's initial vocational preference is assumed to be a powerful method for equating students in the same vocational class across colleges: first, an earlier study (Astin, 1965) revealed that a college freshman's vocational choice and initial major field were the best predictors of his vocational choice at graduation and that other variables added only .01–.03 to the multiple correlation. Second, several analyses were performed to examine the adequacy of initial vocational choice as a control variable. For this purpose, average high school grades and family income were related to initial and final vocational choices in the fall sample for men and women. The median tests (2×6 chi-squares) revealed that a student's family income and average high school grades were significantly related to both his first and last vocational choice. And, for women, high school grades were related more to first than last vocational choice. For men, the converse was true, but the difference was trivial ($\chi^2 = 16.5$ versus 19.7). For women, family income was related more to first than last vocational choice; for men, income was related more to last than first choice, but when the chi-squares were converted to contingency coefficients ($C = .14$ versus $C = .18$) only a slight difference was obtained.

Taken together, these analyses suggest that a student's initial vocational choice is a substantial control that usually incorporates the variance contributed by control variables such as grades and income. In addition, initial vocational choice incorporates and surpasses the contribution that the VPI makes to prediction (see earlier section on prediction).

The institutional environment for each college was characterized by the Environmental Assessment Technique (EAT) (Astin & Holland, 1961). The assignment of major fields to the six environmental models (Realistic, Intellectual, Social, Conventional, Enterprising, and Artistic) was made according to the revised criteria (Holland, 1966b, pp. 122–124). This particular version of the EAT is only slightly different from earlier versions. The source of the basic data was Wright (1966).

The computation of EAT scores was as follows: For each college, the number of degrees earned in each field were assigned, by the criteria indicated earlier, to one of the six environmental models. The absolute numbers for each environmental model were converted to percentages of the total graduating population for that college. The resulting set of six percentages forms the environmental profile for a college. For example, a college of 200 graduating seniors might have the following profile:

Type	*Number*	*EAT score (percent)*
Realistic	20	10
Intellectual	8	4
Social	12	6
Conventional	128	64
Enterprising	28	14
Artistic	4	2
Total	200	100

Stability of Vocational Choice

Several analyses were performed to test the hypothesis that a student is more likely to maintain his vocational choice if he lives in a "congruent" environment—that is, Realistic types will be more stable in Realistic college environments; Intellectual types in Intellectual environments, etc. (Holland, 1966b, p. 73f). First, for each sex at each college, 6×6 tables, based on the classification scheme summarized earlier, were formed to show the relation of first to last vocational choice for time intervals of 8 or 12 mo. for the fall and spring samples. Second, percentages of stability for each vocational class at each college were computed. Third, a total

TABLE 28

CORRELATION (RHO) BETWEEN THE SHAPE OF THE COLLEGE PROFILE (EAT CODE) AND THE STABILITY OF STUDENT VOCATIONAL CHOICES IN THE CORRESPONDING AREAS OF CHOICE

Spring sample

College	Men	Women
A	.03	.35
B	.11	−.35
C	.84*	1.00
D	.64	—
E	.16	1.00*
F	.14	1.00*
G	.20	1.00
H	.41	−.50
I	−.47	−.50
J	.94**	.80
K	−.07	.80
L	.21	.50
M	.24	.88
N	−.54	.80
O	.47	1.00
P	−.03	−.40
Q	.16	1.00
R	.14	.95
S	−.03	.80
T	.26	.80
U	—	1.00
V	—	.13

Fall sample

College	Men	Women
A	.71	.85
B	.03	.65
C	.09	—
D	.20	1.00*
E	.20	.80

Note.—The significance levels for women vary because the number of vocational classes at individual colleges varies from two to four. For men, the number of vocational classes always equals six so that all rhos are based on an N of six.
* $p < .05$.
** $p < .01$.

percentage of stability was calculated for each college, that is, the percentage of *all* students who reported either identical choices or choices that belonged to the same vocational class on both occasions (no changers plus intraclass changers). Fourth, a college's EAT scores were ranked from one through six or from highest to lowest percentage of graduating seniors in each of six areas of study. Finally, the percentages of stability for the six personality types and the six EAT scores were intercorrelated (rho) for each college. The results of these analyses are shown in Table 28; a positive correlation means that high percentages of stability for a given type are associated with high percentages of students graduating in the same area. Conversely, low percentages of stability are associated with low percentages of students graduating in the same area. Put another way, the *positive* correlations suggest that a student is more likely to maintain his present vocational choice if he attends a college in which his choice is congruent with the most popular fields defined according to the present six-category scheme. In contrast, a student is more likely to change fields if he attends a college in which his choice is atypical or unpopular.

The use of a scale of vocational stability (six percentages) and a scale of environments (six percentages) instead of a dichotomy of congruent and incongruent interactions is a comprehensive and severe test of the congruency hypothesis. Many percentages were based on small Ns and are markedly unreliable; the expectation that these two sets of percentages will be ordered in identical fashion is indeed a grandiose hypothesis.

In Table 28, only 5 out of the 50 correlations are significant beyond the 5% level, but 41 out of the 50 are positive, an outcome that is statistically significant by sign test and congruent with the hypothesis. The magnitude of the relationship, however, is ambiguous, for the rhos are based on Ns of only six for men and two to four for women. At best, the results imply that the congruency hypothesis has some tenuous validity.

To secure a more explicit test of the congruency hypothesis, we correlated the percentages of vocational stability for each type with the rank of the corresponding environmental model (EAT variable) across colleges. (The rank of a given EAT variable within a college was reranked across colleges.) Because the between-college analyses in Table 29 are concerned with only a single pair of student-college variables, they are more easily interpreted than the within-college analyses of Table 28, in which each correlation involves a simultaneous test of six pairs of student-

TABLE 29

CORRELATION (RHO) BETWEEN THE RANK OF EAT ENVIRONMENTS AND THE RANK OF THE STABILITY OF CORRESPONDING VOCATIONAL CHOICES

	Spring sample				Fall sample			
	Men	*f*	Women	*f*	Men	*f*	Women	*f*
Real E vs Real VCs	.45*	20	—	—	.67	5	—	—
Int E vs Int VCs	−.01	20	.80**	19	.47	5	−.30	4
Soc E vs Soc VCs	.29	20	−.06	21	.70	5	1.00*	4
Conv E vs Conv VCs	.32	20	.51*	9	.36	5	.74	4
Ent E vs Ent VCs	.01	20	—	—	.86	5	—	—
Art E vs Art VCs	−.33	20	.22	21	−.58	5	.81	4

Note.—*f*'s equal number of *colleges*, not students.
[a] E = environment.
[b] VC = vocational choice.
* $p < .05$, one-tailed test. For N of 20, $p < .05$ = rho of .377.
** $p < .01$.

college variables. Although only 4 out of the 20 correlations (rho) in Table 29 are significant, all 4 are positive in direction. Taken together, all statistically significant findings in Tables 28 and 29 support the congruency hypothesis: students maintain their vocational choices when their college tends to be dominated by students whose vocational choices belong to the same general class.

To test the hypothesis that homogeneous environments are "conducive to vocational stability" (Holland, 1966b, p. 67), the percentage of stability for all students at a college and the homogeneity of a college's environment (the percentage difference between the largest and smallest EAT scores) for samples of males and females in the spring and fall samples were intercorrelated. The correlations (product-moment) in Table 30 are all positive and support the hypothesis, but only one out of the four correlations is statistically significant. The results imply that students are more apt to maintain their original choice when they attend colleges with well-defined or spiked profiles, and that students are more apt to change their vocational aspiration when they attend colleges with flat profiles or colleges with about equal numbers of students in the six vocational classes defined earlier.

The results in Table 30 can also be explained as simply the outcome that would be expected from common knowledge about the size of an educational institution; that is, a student will tend to remain in his field of initial choice in a small college because there are fewer alternatives, and he will tend to change field in a large college because there are many alternatives. To test the validity of this rival hypothesis, stability of vocational choice at an institution was correlated with its size. These results are also shown in Table 30, and they imply that size is a less plausible hypothesis than homogeneity is. The single significant correlation supports the size hypothesis, but the remaining three correlations are not significant and in the wrong direction. In this regard, it should be noted that "homogeneity" is a complex construct that incorporates size as well as variability among environmental variables (EAT); that is, large and diverse institutions are apt to be "heterogeneous" while small institutions are apt to be "homogeneous."

TABLE 30

CORRELATION (PRODUCT-MOMENT) OF STABILITY OF STUDENT VOCATIONAL CHOICE WITH HOMOGENEITY AND SIZE OF AN INSTITUTION

	Spring sample		Fall sample	
	Men (N = 20)	Women (N = 21)	Men (N = 5)	Women (N = 4)
Homogeneity	.22	.60**	.68	.23
Size	.39	.24	.61	−.86*

Note.—Percentages were transformed by arcsin function to render distributions of percentages more normal. N = number of colleges, not students.
* $p < .05$, two-tailed test.
** $p < .01$, two-tailed test.

TABLE 31

THE CORRELATION (RHO) BETWEEN THE SHAPE OF THE COLLEGE PROFILE (EAT) AND INITIAL AND FINAL SATISFACTION WITH COLLEGE

College	Men		Women	
	Initial	Final	Initial	Final
Spring sample				
A	.26	.83*	.79	.39
B	−.39	−.59	.33	.26
C	.13	.30	.13	−.26
D	.26	.67	.64	.86*
E	−.09	.09	−.26	.09
F	−.43	−.03	—	—
G	−.37	−.37	−.37	.04
H	.76	.37	−.20	−.31
I	.20	.33	.14	.26
J	−.31	−.26	.26	.49
K	−.60	−.31	−.66	.33
L	.63	.39	.81	.29
M	.37	.09	.47	.89*
N	.59	.14	−.47	.49
O	−.79	−.19	.53	.56
P	.09	−.37	.44	.56
Q	.03	−.31	.20	.89*
R	−.37	.21	.16	.61
S	.63	.53	.59	−.49
T	—	—	.66	.89*
U	—	—	.43	.31
V	.27	−.81	.44	.62
Fall sample				
A	.43	.31	.21	.71
B	.03	−.47	.43	.60
C	.24	.53	—	—
D	.10	.03	−.71	−.03
E	.03	.81	.90**	.39

* $p < .05$.
** $p < .01$.

Satisfaction with College

To test the hypothesis that a student's satisfaction with his college will be greater if his type is congruent with his college environment (Holland, 1966b, p. 73), initial and final mean satisfaction scores for students of different types were correlated with the ranks of the environmental models of a college. More explicitly, initial and final mean satisfaction scores were computed for the six types at each college. Then the means for the six initial and final satisfaction scores were ranked one through six. Last, ranks for initial and final satisfaction were correlated with the ranks for a college's EAT (one through six). The rhos for these analyses are presented in Table 31. The results for men clearly negate the hypothesis—satisfaction with one's college is associated as much with incongruency of personality types and college environment as it is with congruency.

The results for women appear to support the hypothesis—satisfaction with one's college increases if one's personality type tends to be congruent with the more popular areas of study at the college. For women, 17 out of 25 pairs of correlations for initial and final satisfaction increase over time—trends that suggest congruency of student and college lead to greater satisfaction. Reliance upon the increase in correlation between initial and final satisfaction serves as a useful control for initial level of satisfaction.

In the final analysis, the average final satisfaction score for all students at a college was correlated with the homogeneity score for their college environment. Other things being equal, the results in Table 32 support the hypothesis that students are more likely to be satisfied in colleges characterized by their flat profiles—about equal percentages of personality types. Conversely, students are likely to be less satisfied in colleges with a well-defined or homogeneous student body. These results fit the theory that a student is more likely to find friends if he attends a large, diverse institution (where there will be many students of diverse types) than if he attends

TABLE 32

CORRELATION (PRODUCT-MOMENT) OF HOMOGENEITY AND SIZE OF COLLEGE WITH STUDENT SATISFACTION

	Spring sample		Fall sample	
	Men (N = 20)	Women (N = 21)	Men (N = 5)	Women (N = 4)
Homogeneity	−.54**	−.43*	.00	−.40
Size	.17	.01	−.87**	−.96**

Note.—Percentages were transformed by arcsin function to render distributions of percentages more normal. N = number of colleges, not students.
* $p < .05$.
** $p < .01$.

a small, selective institution (where there will be fewer students of only a few types).

The correlation between institutional size and satisfaction with college is also shown in Table 32. In general, size appears to be a slightly less reliable predictor of satisfaction than homogeneity.

The findings in Tables 31 and 32 imply that a student's satisfaction with his college follows the following theoretical rules:

1. Within the hierarchy of six types of students at a college, the higher a student's choice stands in that hierarchy, the greater the likelihood of satisfaction. This is simply a technical way of saying that students are happier the more closely they resemble the majority of students.

2. Students are more likely to find satisfaction in colleges that possess about equal numbers of each personality type. The assumption here is that heterogeneous colleges have a diverse student population so that a student's opportunity for finding congenial friends is greater than it would be at a college populated largely by a single kind of personality type.

3. The most satisfied student should be a student who resembles the most popular type at his college *and* whose college has a heterogeneous profile of types—the distribution of types is about equal.

4. The most dissatisfied student should be a student who resembles the most unpopular type at his college *and* whose college has a homogeneous distribution of types—the distribution of types is very unequal so that a single type dominates the college with only weak representation of the remaining types. Thus, his chances for finding congenial friends are less than they would be in a heterogeneous student body.

Taken together, the applications of the theory to student-college interactions in Tables 28–32 were a partial success. These analyses, however, clearly reveal the need for replication in a larger number of colleges and for more complete student samples at individual colleges. Further, the findings that contradict the theory require clarification.

DISCUSSION

Although the results were obtained for relatively typical samples of college freshmen and sophomores, the hypotheses tested still need replication. The large sample loss in the follow-up assessments casts some doubt upon the findings despite the failure to find marked differences between the initial and final samples. In addition, the need to generalize the findings to more diverse populations remains, although the typical college student is an improvement over the typical National Merit finalist—the source of data for nearly all earlier investigations.

In this regard, the present study, using typical college students, appears to replicate the major findings obtained earlier with samples of superior scholastic aptitude (Holland, 1962, 1963, 1963–64, 1964; Holland & Nichols, 1964). If anything, the results of the present study are more consistent with the hypotheses in the theory than were the results of earlier studies. For example, students characterized as different types more frequently show the presence of expected traits, and there is less contradictory overlapping of traits than before. Of special importance, it is still possible to distinguish people on the third scale of a three-scale profile *and* this accords with the theory. This severe test of the theory implies that its concepts have considerable explanatory power.

The results of the college-effects analyses in the present and earlier studies are also consistent with one another in that all studies reinforce the congruency hypothesis. But because identical analyses were not performed—largely because they were not possible or appropriate—it is not possible to clearly indicate specific replications or contradictions. This problem is also compounded by the controversy about the best statistical method for the analysis of college influence. The need for replication of these analyses is great, for the findings about institutional influence are tenuous. In addition, large numbers of colleges are required to test all of the main hypotheses in the theory. In the present study, the small sample of colleges precluded the test of many hypotheses and it limited confidence in the outcomes. At the same time,

the general magnitude of the college effects found here are large, perhaps the largest, relative to the institutional influences assessed by others. The results suggest that the theory, in conjunction with the Environmental Assessment Technique (EAT), warrants further examination, for the results imply that the EAT provides a sensitive environmental assessment.

To a limited degree, the results are consistent with a recent study by Astin (1965) in which two out of three positive findings revealed that stability of vocational choice is most closely associated with congruency of a student's vocational choice and the corresponding EAT variable. For instance, Realistic vocational choices were positively correlated with the Realistic Orientation at a college. In another study by Brown (1966), stability of a student's vocational choice was closely related to whether or not he lived in a dorm with students having similar major-field intentions. Brown's experimental study is a precise and unequivocal study of the congruency hypothesis, but congruency was defined in terms of scientific and humanistic types rather than the present types. Similarly, he found that general satisfaction with one's college was positively associated with living in a dorm where the majority of students had similar major-field choices. In his study of vocational plans in college, Davis (1965) also reports substantial evidence for the congruency hypothesis as well as the "birds of a feather" hypothesis. Taken together, these studies support the present theory despite their differences in statistical design and definitions of vocational choice and college environment.

There are several major theoretical implications of the results. First, the definition of types by the classification of a student's vocational choice yields predictions of subsequent choices that are superior to predictions obtained from the definition of types according to the VPI. In future studies, then, it seems advisable to examine the theory by using expressed vocational choice as the main technique for defining a person's type. Similarly, it seems desirable to study "consistency of personality pattern" in terms of a person's first two expressed vocational choices. Consistency of personality pattern might be elaborated as follows: (*a*) a person gives identical responses for his first two preferred vocational choices (when asked for his first two choices, he perseverates and names identical occupations), (*b*) a person gives two choices in the same vocational class, (*c*) a person gives two choices in "consistent" classes, and (*d*) a person gives two choices in "inconsistent" classes. These degrees of consistency promise more positive findings. The substantial predictive efficiency of the analyses of students whose vocational choices belonged to the same or different vocational classes was a partial test of this idea.

Second, the theory needs several revisions. A person's preference for various vocational roles needs to be clarified, for the assignment of roles to types is often contradicted by the evidence. Another more extensive revision of the theory is needed to provide more consistent results in the case of women. The use of the special classification for women improved the prediction of their vocational choices, but many other analyses are still significant for men but not for women.

Third, the success of the VPI as a classifying tool for assigning occupations and fields to classes implies the value of both making the classification more comprehensive and studying more subclasses. The extension of the classification scheme (Holland, 1966a) can be performed simply by assessing representative samples of occupations with the VPI, computing means, and assigning the occupation to the appropriate class in terms of its coded mean VPI profile.

Last, the partial success of the theory as a way to interpret student-institutional interactions implies that employee-organizational interactions and outcomes such as tenure and job satisfaction may follow the same principles. The typological and environmental formulations of the theory would provide a way to control and interpret such interactions.

REFERENCES

ASTIN, A. W. Effect of different college environments on the vocational choices of high aptitude students. *Journal of Counseling Psychology*, 1965, **12**, 28–33.

ASTIN, A. W., & HOLLAND, J. L. The Environmental Assessment Technique: A way to measure college

environments. *Journal of Educational Psychology*, 1961, 52, 308–316.

BROWN, R. D. Manipulation of the environmental press in a college residence hall. Unpublished doctoral dissertation, University of Iowa, 1966.

DAVIS, J. A. *Undergraduate career decisions*. Chicago: Aldine, 1965.

FARBER, I. E., & GOODRICH, L. D. Student Orientation Survey. Preliminary report, Public Health Service research grant M-226, 1964.

FOOTE, N. N., & COTTRELL, L. S. *Identity and interpersonal competence*. Chicago: University of Chicago Press, 1955.

HOLLAND, J. L. A theory of vocational choice. *Journal of Counseling Psychology*, 1959, 6, 35–45.

HOLLAND, J. L. Some explorations of a theory of vocational choice: I. One- and two-year longitudinal studies. *Psychological Monographs*, 1962, 76(26, Whole No. 545).

HOLLAND, J. L. Some explorations of a theory of vocational choice and achievement: II. A four-year prediction study. *Psychological Reports*, 1963, 12, 545–594. Southern Universities Press, 1963, Monograph Supplement 4-V12.

HOLLAND, J. L. Explorations of a theory of vocational choice: IV. Vocational preferences and their relation to occupational images, daydreams and personality. *Vocational Guidance Quarterly*, published in four parts in Summer, Autumn, and Winter issues, 1963–64.

HOLLAND, J. L. *Explorations of a theory of vocational choice: V. A one-year prediction study*. Moravia, New York: Chronicle Guidance Professional Service, 1964.

HOLLAND, J. L. Manual for the Vocational Preference Inventory. (6th rev.) Iowa City, Iowa: Educational Research Associates, 1965.

HOLLAND, J. L. A psychological classification scheme for vocations and major fields. *Journal of Counseling Psychology*, 1966, 13, 278–288. (a)

HOLLAND, J. L. *The psychology of vocational choice: A theory of personality types and model environments*. Waltham, Massachusetts: Blaisdell, 1966. (b)

HOLLAND, J. L., & BAUM, L. L. The Preconscious Activity Scale: The development and validation of an originality measure. *Journal of Creative Behavior*, in press.

HOLLAND, J. L., & LUTZ, S. W. *Predicting a student's vocational choice*. ACT Research Report No. 18. Iowa City: American College Testing Program, 1967.

HOLLAND, J. L., & NICHOLS, R. C. Explorations of a theory of vocational choice: III. A longitudinal study of change in major field of study. *Personnel and Guidance Journal*, 1964, 43, 235–242.

KUBIE, L. S. *Neurotic distortion of the creative process*. Lawrence, Kansas: University of Kansas Press, 1958.

RICHARDS, J. M., JR., HOLLAND, J. L., & LUTZ, S. W. *The prediction of student accomplishment in college*. ACT Research Report No. 13. Iowa City, Iowa: American College Testing Program, 1966.

ROKEACH, M. Political and religious dogmatism, an alternative to the authoritarian personality. *Psychological Monographs*, 1956, 70(18, Whole No. 425).

THOW, M. The campus viewed as a culture. In H. T. Sprague (Ed.), *Research on college students*. Boulder, Colorado: Western Interstate Commission for Higher Education, 1960. Pp. 105–123.

WRIGHT, P. *Earned degrees conferred, 1963–64*. United States Department of Health, Education and Welfare, Office of Education, United States Government Printing Office, 1966.

(Received June 28, 1967)

journal of applied psychology

VOL. 52, NO. 2 APRIL 1968

TEAM PERFORMANCE AS A FUNCTION OF TEAM ARRANGEMENT AND WORK LOAD [1]

WILLIAM A. JOHNSTON AND GEORGE E. BRIGGS

Ohio State University

Pairs of Ss served as radar controllers in a simulated approach-control task, and they were required to alternate in directing aircraft approaches. Prescribed approach rate was 1 approach every min. (high work load) or every 2 min. (low work load). S could compensate for his partner's early or late approaches in compensatory teams but not in noncompensatory teams. Each team completed 4 experimental sessions. Standardized algebraic average of approach times was closer to criterion in compensatory teams than in noncompensatory teams, particularly under low load. Under high load fewer flight errors occurred in compensatory teams than in noncompensatory teams. Team communication inhibited team performance only in the noncompensatory high-load condition. It was concluded that some team functions hinder, others enhance, team output.

The available evidence favors the conclusion that team output is an inverse function of the extent to which teammates are required to coordinate, communicate, or otherwise interact with one another (Briggs & Naylor, 1965; Johnston, 1966; Kidd, 1961; Lanzetta & Roby, 1956, 1957; Naylor & Briggs, 1965; Williges, Johnston, & Briggs, 1966). An important implication of this conclusion is that such team functions should be minimal and individual functions maximal in the design of multiman systems. However, the previous investigations have neglected at least two potentially useful team functions: the *fail-stop* and *compensatory* functions. The former is one in which a team member *prevents* his partner from committing an error, and the latter is one in which a member corrects or counteracts his partner's error *after* it has been committed.

These team functions would seem to be particularly valuable in systems in which errors can eventuate in disaster. For example, one can well imagine the fatal consequences which would ensue if fail-stop and compensatory team functions were not incorporated into surgical and trapeze teams! Of course, the utility of these functions should vary with system load. For example, the frequency of individual errors and therefore the advantage of the fail-stop function should be directly related to the load on the system. On the other hand, the opportunity for compensatory activity should be an inverse function of system load; the higher the load, the more pressed a team member is to complete his own task, and the less time he has to compensate for a teammate's errors. The present study explored the thesis that compensatory and fail-stop team functions can facilitate team output, at least under certain conditions of system load.

In addition, the present study sought to further assess the inhibitory team function of intermember communication (Williges et al., 1966). The effect of this team function should also vary with system load: the greater the load, the less a team can afford the luxury of

[1] This research was carried out at the Human Performance Center and was supported in part by the United States Navy under Contract No. N61339-1327, sponsored by the United States Naval Training Device Center, Orlando, Florida. Reproduction of this publication in whole or in part is permitted for any purpose of the United States Government. The authors are grateful to Robert C. Williges for his help in analyzing the data.

© 1968 by the American Psychological Association, Inc.

team communication, and so the greater should be the disruptive effect of such communication.

METHOD

Task and apparatus. A simulated approach-control task was used in which subjects (*S*s) served as radar controllers (RCs) who were responsible for directing aircraft through an approach gate. The approaches were to be effected at a prescribed rate, and two *S*s were required to alternate in controlling approaches. That is, the first approach was to be directed by RC_1, the second by RC_2, and then RC_1 assumed responsibility for the third incoming aircraft, and so on.

The radar displays were linked to target-generator consoles via a special-purpose analog computer (Hixson, Harter, Warren, & Cowan, 1954). The role of pilots was assumed by well-trained experimental assistants who worked at the target-generator consoles. The RCs delivered instructions to pilots over a voice-communication channel, and the pilots carried out the instructions faithfully and without delay by appropriate manipulation of their consoles. The RC-to-pilot communication protocol required that the RC first identify the pilot and then issue the command, for example, "Bravo one, speed 200 knots"; in return, the pilot was to immediately confirm the command, for example, "Roger; Bravo one, speed 200 knots." The RCs gave only heading and speed commands to the pilots; altitude was intentionally omitted from consideration in order to maintain a reasonable level of task difficulty.

One of the displays as it appeared at the beginning of a session is reproduced in Figure 1. The approach gate was located precisely at the center of the displayed airspace, and all aircraft entered the airspace from the eastern (right-hand) periphery. The display was marked with concentric rings, and the distance between adjacent rings represented 20 mi. The distance from the periphery to the approach gate represented 100 mi. The aircraft appearing in the northeast sector of the scope were designated as "Alpha" aircraft, and those appearing in the southeast sector were referred to as "Bravo" aircraft. The Alpha and Bravo aircraft were indicated by different codes. The two RCs monitored the same airspace but on different displays, and they could speak to one another only over a voice-communication channel. The *Alpha* RC was assigned to the Alpha aircraft, the *Bravo* RC to the Bravo aircraft. The team-communication protocol required that both RCs identify themselves before each conversation. All team communications were tape-recorded. A partition was erected between the two RCs so they could not see one another, and an ambient noise generator prohibited the RCs from hearing one another except by means of the inter-RC communication link.

At the beginning of a session, four Alpha aircraft and four Bravo aircraft were spaced evenly along the eastern periphery of the airspace. An

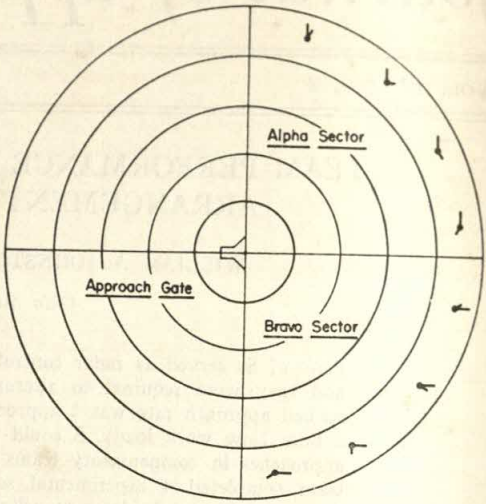


FIG. 1. Reproduction of a display at the start of a session.

Alpha aircraft made the first approach, a Bravo aircraft followed, then another Alpha aircraft, and so on. A successful approach occurred when an aircraft entered the approach gate at 200 kn. on a heading of 270°. A "miss" occurred if an aircraft in the inner circle of the display crossed the longitudinal axis into the western half of the airspace in any condition not constituting a successful approach. Aircraft were removed from the display following successful landings or misses, and as soon as one aircraft was removed a new one was positioned at the eastern periphery. A safety infraction occurred if two aircraft came within 2 mi. of one another. However, a safety infraction did not lead to the removal of the aircraft.

The entire operation, including the scoring and recording of approaches, misses, and safety infractions, the presentation and removal of aircraft, and the administration of feedback, was conducted by a team of two experimental assistants.

Subjects and design. The *S*s were 64 male undergraduates who served in pairs for five consecutive sessions. All were paid for their services at the rate of $1.25 per session. Four conditions were formed from the factorial combination of two levels of team arrangement (compensatory and noncompensatory) with two levels of system load (low and high). Teams were composed of *S*s who had volunteered for the same experimental time period, and they were randomly assigned to the four conditions (groups) with the restriction that groups be filled equally across the 8 wk. of data collection.

Procedure. The first session was devoted to general instructions and practice, and the experimental conditions were implemented during the remaining four sessions. Each session required 35 min., and the intersession interval was 24 hr.

The required approach rate (*system criterion*) was an approach every minute for high load, and

one every 2 min. for low load. In the noncompensatory arrangement, the system criterion was to be satisfied on *each* approach. In the low-load condition, for example, a given approach was to be effected precisely 2 min. after the preceding aircraft passed the gate regardless of any time error that might have accrued over prior approaches. By contrast, a given approach in the compensatory arrangement was to compensate for the accumulated time error. Hence, if a time error of 20 sec. late had accrued over prior approaches, the next approach was to be 20 sec. early so that the *average* of the approach times would equal the system criterion.

A clock system was mounted on the wall in front of the team in such a way that the RCs could keep continuous track of their temporal progress in guiding aircraft through the approach gate. Each clock kept time in terms of minutes and seconds up to 1 hr., and could be viewed by only a slight shift in an RC's line of vision from the input display. Only one clock was used in the compensatory condition, and the RC team was instructed to have an aircraft through the approach gate at every one (high load) or two (low load) complete rotations of the second hand. In order to avoid anomalies that arise when unusually large time errors accrue over previous landings, the clock was restarted after every six approaches.

Two clocks were used in the noncompensatory condition: one for the Alpha RC, the other for the Bravo RC. When the Alpha RC made an approach (regardless of how early or late the approach might have been), the Bravo RC's clock would start and a Bravo aircraft was to be at the approach gate after either one (high load) or two (low load) complete rotations of the second hand. Then, when the Bravo aircraft passed through the approach gate, the Alpha RC's clock (which had been reset during the Bravo approach) would start again, and so on.

The following aspects of the task characterized both team arrangements: (*a*) A small red light indicated whose turn it was to effect an approach; (*b*) the timing started upon the completion of the first approach of the session; and (*c*) immediate feedback was provided to the RC team of time errors (relative to the compensatory or noncompensatory criteria), misses, and safety infractions immediately after each approach except the first.

Performance measures. The *experimental* timing of each approach began when the immediately preceding aircraft passed through the approach gate. Approach-time *error* was defined as the difference between actual approach time and desired *average* approach time (actual minus 60 sec. or 120 sec. depending on the system criterion in effect). Thus, a +20-sec. time error indicated an approach that was 20 sec. late (e.g., an approach time of either 80 sec. or 140 sec.). Both the algebraic and the absolute sums of the time errors were calculated for each team on each of the last four sessions. The algebraic or constant error (*CE*) score was then divided by the absolute or variable error (*VE*) score to provide a single measure for each team on each session.

The *CE/VE* ratio is analogous to the *z* or standard score; thus, it is a way to compare *CE*s from populations differing in variability. In the present study the *CE/VE* ratio has special implications for the compensatory versus noncompensatory comparisons. In effect, the instructions to the compensatory teams emphasized small *CE*s (at the expense of *VE*), while the instructions to the noncompensatory teams, in effect, emphasized small *VE*s (at the expense of *CE*). Relative to the noncompensatory arrangement, therefore, the compensatory arrangement was expected to reduce *CE* but to elevate *VE*. In short, smaller ratio scores were expected under the compensatory condition than under the noncompensatory arrangement. Further, this difference was expected to be particularly pronounced under low system load because this condition provides minimal constraints on compensatory activities.

Each team was scored also in terms of total number of flight errors (misses and safety infractions) per session. It was anticipated that an RC would pay more attention to his partner's activity, and would therefore be more likely to prevent his partner from committing errors, in the compensatory arrangement than in the noncompensatory arrangement. This fail-stop function should be most pronounced under high load, the condition in which flight error frequency should be maximal. Hence, error frequency was expected to be higher in the noncompensatory arrangement than in the compensatory arrangement, particularly under high load.

Team-communication measures. All team communications were categorized according to the content-analysis scheme described elsewhere (Williges et al., 1966). This scheme is based on five major categories of communication: *identification, irrelevant, declarative, tactical,* and *command* messages. Teammates should have more reason to communicate with one another in the compensatory arrangement than in the noncompensatory arrangement, and should have more "freedom" to do so under low-load than under high-load conditions. It follows that team communication should be most frequent in the compensatory low-load condition and should be most disruptive in the noncompensatory high-load condition.

RESULTS

Team performance. An analysis of variance of *CE/VE* ratios produced the following significant effects: system load, $F(1/28) = 12.63$, $p < .01$; team arrangement, $F(1/28) = 4.24$, $p < .05$; sessions, $F(3/84) = 6.49$, $p < .01$; and the System Load × Team Arrangement × Sessions interaction, $F(3/84) = 4.56$, $p < .01$. The character of these effects

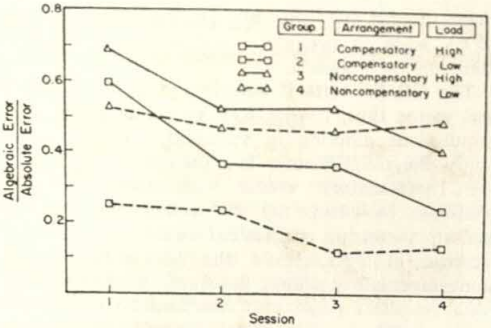


FIG. 2. The ratio of algebraic error to absolute error for each experimental condition as a function of sessions.

is depicted in Figure 2. Specifically, the *CE/VE* ratio decreased across sessions, increased with load, and was higher in noncompensatory teams than in compensatory teams, particularly on Session 1 under low system load. In accordance with prediction, therefore, the extent to which "standardized" algebraic error was reduced at the expense of absolute error was greatest in the compensatory low-load condition. The sessions effect arose from a greater intersession decline in algebraic error than in absolute error, which finding suggests that practice benefited *CE* more than it did *VE*.

The time error per approach averaged 11.13 sec. and 18.43 sec. under low and high load, respectively. These data indicate that the RC teams were under considerable time stress even under "low" load. Had the low-load criterion been, say, one approach every 3 or 4 min., then the effect of team arrangement on the *CE/VE* ratio may have been even more obvious.

The following sources of variance were significant in the analysis of variance of frequency of flight errors: load, $F(1/28) = 13.01$, $p < .01$; sessions, $F(3/84) = 55.81$, $p < .001$; Load × Sessions, $F(3/84) = 3.49$, $p < .05$; and the Load × Team Arrangement × Sessions interaction, $F(3/84) = 3.33$, $p < .05$. The nature of these effects is apparent in Figure 3. The frequency of flight errors increased with system load and decreased across sessions. As was expected, flight errors were less frequent in the compensatory arrangement than in the noncompensatory arrangement, but only in the first experimental session under high load.

Team communication. Team arrangement, $F(1/28) = 10.86$, $p < .01$, communication category, $F(4/112) = 26.72$, $p < .001$, and Team Arrangement × Communication Category, $F(4/112) = 2.83$, $p < .05$, were all that attained significance in the analysis of variance of communication frequency. Contrary to expectation, team communication was not influenced by load. In agreement with expectation, however, communication frequency was higher in compensatory teams (an average of 105 per session) than in noncompensatory teams (47 per session). Table 1 shows the relative frequencies of the five communication categories for each team arrangement. The relative frequencies of declarative messages, tactical communications, and commands were higher in compensatory teams than in noncompensatory teams, and the relative frequency of irrelevant communications was higher in noncompensatory teams than in compensatory teams. Identifications were relatively frequent in both team arrangements.

Correlational analyses were performed to assess the relationship between communication frequency and team performance (all measures averaged across sessions). Significant correlations were obtained only with respect to the noncompensatory high-load condition, in which Condition *CE/VE*, $r(6) = .81$, $p < .05$, *CE* (algebraic error), $r(6) = .93$, $p < .01$, and *VE* (absolute error), $r(6) = .91$, $p < .01$, were directly related to communication frequency. It may be recalled

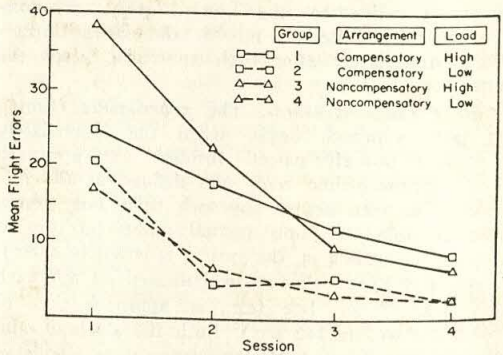


FIG. 3. The total frequency of flight errors for each experimental condition as a function of sessions.

TABLE 1

MEAN RELATIVE FREQUENCY OF OCCURRENCE OF EACH COMMUNICATION CATEGORY FOR BOTH TEAM ARRANGEMENTS

Communication category	Team arrangement	
	Compensatory	Noncompensatory
Identification	.28	.32
Irrelevant	.06	.16
Declarative	.27	.22
Tactical	.20	.15
Command	.19	.15

that noncompensatory teams were instructed, in effect, to maximize *CE/VE* by reducing *VE* at the expense of *CE*. At first glance, it might appear that, under high work load, team communication aided noncompensatory teams in their effort to maximize *CE/VE*. However, team communication augmented *CE/VE not* by reducing *VE* and elevating *CE*, but rather by increasing *both* error indexes, the increase apparently being more pronounced for *CE* than for *VE*. In actuality, therefore, team communication disrupted the performance of noncompensatory teams under high load. That this disruptive effect of team communication was restricted to the noncompensatory high-load condition is consonant with prediction and with the fact that irrelevant communications were relatively prominent in the communication profile of noncompensatory teams.

DISCUSSION

The present data lend testimony to the proposition that team performance can be augmented by the compensatory and fail-stop team functions. Despite the fact that the absolute level of system load was quite high even in the low-load condition, compensatory teams were able to partially offset accrued time error. Thus, the compensatory arrangement would be of service in real approach-control teams when the normal flow of air traffic has been impeded as a result of, say, adverse weather. However, this value of compensatory activity appears to be an inverse function of system load. On the other hand, the utility of the fail-stop function is apparently a direct function of system load. Consequently, when system load is high, a team member may not be able to undo his partner's errors once they have been committed, but he may be able to reduce the number of errors that his partner in fact does commit. Hence, where one team function loses value, the other gains.

Turning now to the communication data, it is apparent that team-communication frequency was augmented by compensatory activity but was not affected by system load. However, it is again noteworthy that the absolute level of work load was quite high even in the so-called low-load condition; a greater disparity in the two levels of load might have induced a disparity in communication frequency. In regard to the role of team communication in teamwork, the present findings further substantiate the previously found disruptive effect of team communication on team performance (Briggs & Naylor, 1965; Johnston, 1966; Williges et al., 1966). Of course, since this disruptive effect was limited to the noncompensatory high-load condition, it appears that team communication hinders team performance the most when there is the least need and the least freedom to communicate. When there is enough time and perhaps some reason to communicate, team communication neither retards *nor* enhances team performance.

In summary, certain team functions such as team communication are potentially detrimental to team output and should be restricted in system design and by the operating procedures. Contrariwise, other team functions such as compensation and error prevention are potentially advantageous and should be incorporated into the design of those systems for which the unique features and effects of these team functions are appropriate.

REFERENCES

BRIGGS, G. E., & NAYLOR, J. C. Team versus individual training, training task fidelity, and task organization effects on transfer performance by three-man teams. *Journal of Applied Psychology*, 1965, 49, 387–392.

HIXSON, W. C., HARTER, G. A., WARREN, C. E., & COWAN, J. D., JR. An electronic radar target simulator for air traffic control studies. *USAF*

WADC Technical Report, December 1954, No. 54-569.

JOHNSTON, W. A. Transfer of team skills as a function of type of training. *Journal of Applied Psychology*, 1966, **50**, 102-108.

KIDD, J. S. A comparison of one-, two-, and three-man work units under various conditions of work load. *Journal of Applied Psychology*, 1961, **45**, 195-200.

LANZETTA, J. T., & ROBY, T. B. Group performance as a function of work distribution patterns and task load. *Sociometry*, 1956, **19**, 95-104.

LANZETTA, J. T., & ROBY, T. B. Group learning and communication as a function of task and structure "demands." *Journal of Abnormal and Social Psychology*, 1957, **55**, 121-131.

NAYLOR, J. C., & BRIGGS, G. E. Team training effectiveness under various conditions. *Journal of Applied Psychology*, 1965, **49**, 223-229.

WILLIGES, R. C., JOHNSTON, W. A., & BRIGGS, G. E. Role of verbal communication in teamwork. *Journal of Applied Psychology*, 1966, **50**, 473-478.

(Received April 19, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 2, 95–97

LEARNING OF TELEPHONE NUMBERS WITH AND WITHOUT LETTER-EXCHANGE PREFIXES

SHELDON J. LACHMAN

Wayne State University

In a group situation, 36 college students, randomly assigned to 4 groups, practiced and were tested for recall on 8- or 10-unit lists of telephone numbers with letter exchanges or with digits substituted for the letters. In terms of mean scores and final achievement levels, groups practicing lists of telephone numbers with lettered exchanges performed better than their counterparts who practiced corresponding lists of telephone numbers composed entirely of digits. Items in the traditional telephone identification system employing pairs of prefixed letter exchanges were most readily learned. Explanations in terms of (a) symbol isolation, (b) interference, and (c) familiarity effects are indicated; implications for telephone identification are specified.

Number-exchange systems in recent years have substantially replaced letter-exchange systems in telephone numbers. Certain individuals and even protest organizations, for example, the Anti-Digit Dialing League, have expressed opposition to telephone company policy in that regard. In part, their opposition has been based on the alleged increased difficulty in learning an all-digit code as opposed to learning a letter-digit code.

Inquiry into the research literature did not yield reports of experiments in the area of telephone number learning. Surprisingly, requests for such research reports of Bell Telephone Laboratories also yielded no experimental reports nor citations of any reports.

The purpose of the study now reported was to obtain evidence concerning the relative efficiency of learning lists of telephone numbers having alphabetical or letter exchanges—two letters and five digits—as contrasted with learning telephone numbers composed of digits exclusively—seven digits.

METHOD

Subjects. The *Ss* were 36 college students, most of them sophomores. They were assigned at random into four groups, nine *Ss* in each group.

Procedure. Group A practiced a 10-unit list; each unit was composed of 2 letters and 5 digits. This is the letter-exchange list. Group B practiced a 10-unit list; each unit was composed of 7 digits. This is the all-digit list. Group C practiced an 8-unit list, each unit composed of 2 letters and 5 digits. These were the first 8 items in the list used by Group A *Ss*. Group D practiced an 8-unit list, each unit composed of 7 digits. These were the first 8 items in the list used by Group B *Ss*.

Material. Learning lists were prepared by an extensive sampling of the *1963 Chicago Telephone Directory* (white pages). About 500 letter-exchange items were selected. From the pool of items, a learning list was developed. The attempt was made to obtain, for the learning list, exchanges which were different from those used in the Detroit area, where this study was conducted; this was not completely successful. However, no set of exchanges was repeated in the list. No common letters were assigned to successive units. As a matter of fact, insofar as possible—again not completely successful—the attempt was made to avoid the use of any letter more than once in the series. Excluded were digit combinations which would make a word such as BE, HE, SO, AT, and IN. By such exclusion, it was believed that difficulty level for learning letter exchanges would be increased since nonword letter combinations would presumably be less familiar. Because there was the possibility of confounding O as a letter with 0 as a number, that symbol was also excluded.

Materials were presented to *Ss* in the form in which they were found in the telephone directory. An equivalent all-digit list was transmuted from the letter-digit list. The learning materials are indicated below:

Letter-Exchange List	All-Digit List
TR 4-9153	874-9153
KE 6-1748	536-1748
WH 3-1826	943-1826
MU 5-9214	685-9214
VA 6-1394	826-1394
PL 2-6843	752-6843
FR 6-2419	376-2419
SA 1-8935	721-8935
MI 2-1569	642-1569
HU 3-2965	483-2965

The *Ss* performed under group conditions. Each *S* had a copy of a duplicated list.[1] They worked at long tables and had abundant room for materials and for writing. The *Ss* were given 20 practice periods each of which was immediately followed by a recall period. During the recall period, responses were written on response forms coded from 1 to 20; on completion of the recall period, response forms for each *S* were deposited by each *S* into *S*-coded envelopes. Each practice period was 30 sec. long and each recall period was 60 sec. long. Intervals between the practice-recall sequences were about 10 sec. in duration.

RESULTS

All groups and individual *Ss* made progress in the direction of learning. Groups A, B, C, and D obtained mean scores of 4.8, 3.5, 4.8, and 3.3, respectively. In each case, the groups practicing telephone numbers with letter exchanges performed better than those having lists of the same length with appropriate numbers substituted for the exchange letters.

An analysis of learning progress by grouping the data in terms of deciles, quintiles, and quartiles disclosed that at each point in the learning sequence the letter-digit group is superior to its parallel all-digit group. Performance of Groups A, B, C, and D on the twentieth, that is, the final, trial were 7.3, 6.0, 6.9, and 5.9, respectively.

Individual differences in performance were obvious. In each group, with regard to total performance on the 20 trials, the score of the best performer was more than double that of the poorest performer.

DISCUSSION

Despite "loading against" letter-digit combinations (by excluding two-letter word exchanges which would ordinarily occur sometimes by chance), results indicate that from the standpoint of facility in learning telephone numbers, the traditional system with a letter exchange is more efficient than the "exchangeless" or all-digit system. With a 10-unit list, on the average, about 37% more material is learned of the conventional letter-exchange phone numbers over comparable all-digit phone numbers. With an 8-unit list, on the average, about 45% more material is learned of the conventional letter-exchange phone numbers over comparable all-digit phone numbers.

Two related viewpoints may be pertinent in attempting to explain the major finding of this research, namely, the superiority of letter-exchange learning. These are termed the isolation viewpoint and the interference viewpoint.

1. The isolation viewpoint maintains that a small group of symbols of one kind among a more numerous group of symbols of another kind are apprehended and learned more readily, perhaps as a function of contrast. In any case, this is related to Von Restorff's (1933) finding that if a syllable is inserted in a list of numbers or a number in a series of syllables the exceptional or "isolated" item will be better recalled.

2. The interference viewpoint maintains that because of the similarity or relative homogeneity of items in the all-digit list, proactive and retroactive influences within the list may have interfering effects on recall. A variety of researches on proactive and retroactive effects provide research support for such a position; much of the literature in this area has been summarized by McGeoch and Irion (1952, pp. 404–447) and by Underwood (1957).

There is also a third possibility which is that letters are easier to learn than numbers; this may be a function of greater frequency of contact (experience) with letter material than number material under ordinary circumstances.

Assuming that the results are meaningful and generalizable, for the convenience of their subscribers, telephone companies might consider a return to the earlier more conventional letter-exchange system involving, if not a name exchange, at least a pair of letters. The use of letters as opposed to numbers provides for a distinct means of differentiating, designating, and identifying—perhaps, in most cases, with less opportunity for retroactive and proactive inhibition. Further, one might have several different letter designations for the same set of numbers. For example, a 74 could be PI, RI, or SI as well as

[1] The paired-associates technique was considered as an alternative method, but several problems concerned with that method and the primary objective of this study suggested that the serial method was more desirable.

PH, RH, and SH. This would achieve better differentiation than using 74 each time. 749-1752 and 748-6945 might be better learned as SI 9-1752 and PH 8-4645. This, of course, assumes that in the lists to be learned, other combinations than 74 are most numerous.

The letter exchange need not necessarily be the first two letters of a meaningful word as has usually been the case. For example, the 83 "exchange" might be dialed if the prefixes were VD, TD, or UF, none of which are words or word prefixes in the English language; but all are distinct from each other and all are commonly used abbreviations and, therefore, not completely without meaning.

Three-letter combinations of words or of nonsense syllables are also not beyond the realm of possibility.

Unfortunately, except for research involving digit span—long a standard method for assessing rote memory—relatively little research has been done with the learning of number material. Recently, however, Battig and Spera (1962) have reported association values for numbers from 0 to 100.

REFERENCES

BATTIG, W. F., & SPERA, A. J. Rated association values of numbers from 0 to 100. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 200–202.

MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning*. New York: Longmans, 1952.

UNDERWOOD, B. J. Interference and forgetting. *Psychological Review*, 1957, **64**, 49–60.

VON RESTORFF, H. Über die Wirkung von Bereichsbildungen im Spurenfeld. In W. Köhler & H. von Restorff (Eds.), Analyse von Vorgängen im Spurenfeld. 1. *Psychologische Forschung*, 1933, **18**, 299–342.

(Early publication received July 18, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 2, 98–103

PREDICTION OF SALES SUCCESS FROM FACTORIALLY DETERMINED DIMENSIONS OF PERSONAL BACKGROUND DATA

MELANY E. BAEHR AND GLENN B. WILLIAMS

Industrial Relations Center, University of Chicago

A concurrent validity study was made of the scores of 210 salesmen and 16 district managers on 15 personal background dimensions, identified previously by factor analysis. Differences between factor-score means for the manager and sales groups and for the upper- and lower-rated sales groups were significant for the factors Financial Responsibility (.001, .001), Early Family Responsibility (.01, .01), and Stability (.001, .05). Multiple-regression analyses of the factor scores against each of 5 criterion measures of on-the-job behavior yielded correlations of .42, .50, and .36 for the 3 major criteria of overall performance, and mean and maximum sales volume rank. The criteria route difficulty and tenure gave correlations of .27 and .30. Interpretation of the highest-weighted factors in these analyses indicates the logical, dynamic relationships between personal background and job behavior.

The bulk of research on the prediction of occupational success from quantitative personal-background-data items has taken an essentially empirical approach (Buel, 1965; Dunnette & Maetzold, 1955; Lockwood & Parsons, 1960; Minor, 1958; Mosel & Wade, 1951; Scollay, 1956; Smith, Albright, & Glennon, 1961). It first determines the discriminative power of individual items in measuring a given criterion of success in some occupation. Then it combines those items which discriminate (with or without weighting) into an instrument which yields an index or "score." Such scores have been individually cross-validated for a variety of occupations in industry and the armed services. They have also contributed materially to multiple correlations of test batteries with empirical criteria of performance. However, the approach has been criticized for its basic empiricism (Dunnette, 1962; Morrison, Owens, Glennon, & Albright, 1962) and its failure to shed any light on the dynamic relationships between personal background on the one hand and occupational choice and success on the other.

The present investigation is a sequel to a study designed to identify and define some of the underlying dimensions of personal background data (Baehr & Williams, 1967). The dimensions were identified through successive factor analyses of responses to a wide variety of background items by an occupationally heterogeneous sample of 680 male employees. The final factor analysis yielded 15 interpretable oblique first-order factors. A factoring of the correlations among these primaries produced five uncorrelated second-order factors which appeared to represent broad behavior patterns associated with the needs and achievement drives of individuals.

Two tests were run on the efficacy of the background dimensions, as measured by the first-order factor scores, in differentiating among the 10 occupational groups covered by the sample. An analysis of variance yielded F ratios which were virtually all significant beyond the .05 level of confidence, the majority of them beyond the .001 level. A complete series of t tests revealed that each pair of occupational groups showed differences which were significant at the .05 level or better and, conversely, that each factor discriminated significantly among some of the pairs.

These findings were considered an adequate basis for concluding that the dimensions of the personal background data were related to occupational classification. Further, assuming that an individual's occupational classification is to some extent under his own control, then one can assume that the background dimensions are also related to occupational choice. Moreover, since the ma-

jority of the cases in the sample had been in their given occupation for some years—and thus must be regarded as at least moderately successful in it—the final assumption can be made that the background dimensions are related to occupational success. In the present study, this final hypothesis was experimentally investigated. The Ss were the 210-man sales force and its 16 district managers of a specialty-food manufacturing organization with branches at various locations throughout the United States.

METHOD

Variables

The predictor variables were the unit-weight factor scores derived from the 15 first-order factors. These factors are defined as follows:

Factor 1 (*School Achievement*). Academic achievement, particularly in high school, but also, where applicable, at college. A general liking for, and adjustment to, the school environment.

Factor 2 (*Higher Educational Achievement*). Special or technical accomplishment and qualification resulting in a relatively late vocational start and late assumption of family responsibilities. This pattern is characterized by eventual occupational and financial achievement.

Factor 3 (*Drive*). Inner drive to be outstanding in performance, to attain high goals even if this entails temporary setbacks, to supervise others, and to achieve success and advancement.

Factor 4 (*Leadership and Group Participation*). A desire to establish contact with others as shown by membership and interpersonal activity in organizations and an interest in influencing others through community and social activities.

Factor 5 (*Financial Responsibility*). Ability to manage a personal economy of defined proportions—to earn, invest, save, and accumulate.

Factor 6 (*Early Family Responsibility*). Early marriage and establishment of a family, with the husband ordinarily being the sole provider. Demonstrated achievement in handling the family's financial affairs. Outside the work situation, the greatest interest is in family activities.

Factor 7 (*Parental Family Adjustment*). Development of realistic, constructive attitudes and relationships in the early family environment. Included are relationships among siblings, between parents, and between the child and the parents.

Factor 8 (*Stability*). Established security and stability in the work situation, resulting from a past history of good performance. The present concern is more to maintain what has been achieved than to plan for improvement or development.

Factor 9 (*School Activities*). Major emphasis on active participation in athletic and extracurricular social activities at high school but also indications of good academic achievement.

Factor 10 (*Professional Successful Parents*). A parental background characterized by a successful father, either self-employed or in one of the professions, and by material comfort and a happy home life.

Factor 11 (*Educational-Vocational Consistency*). A preference for occupations which are highly related (or similar) and are in line with educational interests and training.

Factor 12 (*Vocational Decisiveness*). Decisiveness in choosing an occupation and purposefulness in achieving the necessary qualifications, followed by an early start in the chosen occupation.

Factor 13 (*Vocational Satisfaction*). Satisfaction with occupational choice and the expectation that peak performance will be some time in the future. There is, however, no evidence of consistency of application or of drive to achieve high standards of performance.

Factor 14 (*Selling Experience*). Various kinds of selling experience, including door-to-door selling and transactions in real estate.

Factor 15 (*General Health*). Generally better than average health over an extended period (childhood, adolescence, adulthood). General freedom from physical ailments and from loss of work time resulting from illness.

The criterion indexes were five measures of on-the-job performance. The three major ones of these were directly related to success in selling.

Paired-comparison performance rating. Three judges—ordinarily the district manager, the branch manager, and either the office manager or the buyer for the district—rated the salesmen for overall performance, using the paired-comparison technique. Their preferences were converted to normalized standard scores and then averaged to yield the composite performance measures used in this investigation.

Mean sales volume rank.[1] For the past 10 yr. the organization has ranked each member of its sales force on the basis of his volume of sales for the year. The criterion measure used here is the average of all ranks assigned to a salesman over the period in which his sales volume was ranked.

Maximum sales volume rank. (See Footnote 1.) This figure is the highest ranking the man received over the last 10 yr. or over his period of tenure if this were less than 10 yr.

Route difficulty. Each branch manager used the paired-comparison technique to rate the sales routes in his region from most to least difficult in the

[1] In the calculations presented in this paper, the ranks have been reflected. Thus, a high rank indicates a high sales volume. The commission received by salesmen is based directly on sales volume. For this reason, sales volume was selected as a performance criterion rather than pay, where the performance aspect is obscured by higher base pay for long-tenure salesmen.

TABLE 1

CORRELATIONS BETWEEN CRITERION MEASURES

Criterion measure	Performance rating	Mean sales rank	Maximum sales rank	Route difficulty	Tenure
Performance rating		.37***	.24***	−.20**	−.13
Mean sales rank	.37***		.75***	.02	.18**
Maximum sales rank	.24***	.75***		.02	.16*
Route difficulty	−.20**	.02	.02		−.02
Tenure	−.13	.18**	.16*	−.02	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

realization of sales potential. These ratings were converted to normalized standard scores.

Tenure as a salesman. This measure was the number of years the individual had been employed in the organization as a salesman.

The correlations between these criterion measures are shown in Table 1.

Subjects

The sales personnel who were *S*s in this study were not part of the sample used in identifying the personal background dimensions. The sales groups in the original 680-man sample were paper-forms salesmen and district sales managers in the insurance business. The 210 salesmen in the present investigation were employed in a food-products manufacturing organization supplying institutions and retail outlets. Their average education was 1–2 yr. of college and their average age 42 yr.

The 16 district managers, representing the first line of sales management in the organization, had about the same educational background as the salesmen. Their average age was 48 yr. Since the organization has a strong policy of promotion from within and since the district manager's job still includes a fair amount of personal selling, it may be assumed that the district managers represent the upper end of the performance continuum for the sales force.

Procedure

The *S*s' factor scores were secured by administering a multiple-choice questionnaire called the Personal History Index (Baehr, Burns, & McMurry, 1965).

T-test analyses. A *t*-test analysis was made of the significance of the differences between the mean personal-history factor scores of the group of district managers and of the sample sales group. Another such analysis was run on the differences between the mean factor scores of those salesmen with ratings of 55 and above on the standard-score scale for the paired-comparison assessments and those with ratings of 45 and below.

Multiple-regression analyses. A multiple correlation for the personal background factors was calculated separately for each of the five criterion measures.

RESULTS

Results of the *t*-test analyses are shown in Tables 2 and 3. The three factors of Financial Responsibility, Early Family Responsibility, and Stability differentiated at statistically significant levels both between the salesman and district manager groups and between the upper- and lower-rated sales groups. The scores on these factors rise successively from the lower-rated salesmen through the upper-rated salesmen to the district managers. These results correspond with the assumed level of sales performance of each of these groups. It is interesting to note that the three differentiating factors are among the first 8 of the 15—those which had previously been identified as having the most potential for operational use (Baehr & Williams, 1967).

Results of the multiple-regression analyses are given in Table 4. The computer programs for these analyses were written in such a way that a variable would be included only if its degree of variance produced an *F* ratio at the .05 level of confidence or better. The objectives of this programming were to prevent undue "forcing" of variables into the regression equations and to lessen the effects of purely chance relationships of the variable and the criterion.

The three major criterion measures of success are the paired-comparison performance rating, the mean sales volume rank, and the maximum sales volume rank. Table 1 shows that the latter two measures (because of their common source data) are highly correlated

TABLE 2

SIGNIFICANCE OF THE DIFFERENCES BETWEEN MEAN PERSONAL-HISTORY FACTOR SCORES FOR THE DISTRICT SALES MANAGER AND THE SALESMAN GROUPS

Personal-history factor	District sales managers ($N = 16$)		Salesmen ($N = 210$)		t
	M	SD	M	SD	
School Achievement	2.44	1.37	2.19	1.56	.62
Higher Educational Achievement	5.37	1.32	4.58	1.79	1.75
Drive	4.75	2.22	4.23	2.00	.99
Leadership and Group Participation	2.81	1.67	4.78	1.67	.10
Financial Responsibility	6.50	1.06	4.86	1.80	3.57**
Early Family Responsibility	9.56	1.54	7.95	2.43	2.60*
Parental Family Adjustment	7.00	1.73	6.62	2.23	.66
Stability	7.31	.77	5.56	2.11	3.29**
School Activities	3.31	2.02	3.28	1.87	.26
Professional Successful Parents	4.19	1.38	3.73	1.58	1.13
Educational-Vocational Consistency	1.69	.98	1.33	.97	1.40
Vocational Decisiveness	2.75	1.15	2.17	1.17	1.91
Vocational Satisfaction	6.12	1.69	5.80	1.80	.70
Selling Experience	6.56	1.54	6.23	1.48	.85
General Health	3.19	1.55	3.44	1.08	.86

* $p < .01$.
** $p < .001$.

($r = .75$). The performance rating shows moderate correlations ($r = .37$ and .24) with the mean and the maximum sales volume ranks.

The multiple correlation for the performance-rating criterion was .42 and for the mean and maximum sales volume ranks, .50 and .36, respectively. In the regression anal-

TABLE 3

SIGNIFICANCE OF THE DIFFERENCES BETWEEN MEAN PERSONAL-HISTORY FACTOR SCORES FOR THE UPPER-RATED AND THE LOWER-RATED SALESMEN ON THE PAIRED-COMPARISON PERFORMANCE INDICES

Personal-history factor	Upper-rated salesmen ($N = 72$)		Lower-rated salesmen ($N = 52$)		t
	M	SD	M	SD	
School Achievement	1.96	1.52	2.19	1.65	.82
Higher Educational Achievement	4.59	1.68	4.69	1.91	.30
Drive	4.13	2.04	4.15	1.90	.00
Leadership and Group Participation	2.90	1.66	2.44	1.78	1.48
Financial Responsibility	5.54	1.43	4.11	1.79	4.92***
Early Family Responsibility	8.63	1.98	7.31	2.63	3.17**
Parental Family Adjustment	6.47	2.25	6.79	2.06	.80
Stability	6.01	1.68	5.08	2.56	2.46*
School Activities	3.32	1.88	3.08	1.76	.73
Professional Successful Parents	3.79	1.46	3.60	1.55	.73
Educational-Vocational Consistency	1.42	.95	1.25	.99	.95
Vocational Decisiveness	2.22	1.35	2.14	1.12	.39
Vocational Satisfaction	5.69	1.81	5.67	1.72	.00
Selling Experience	6.25	1.39	5.92	1.44	1.27
General Health	3.42	1.15	3.27	1.11	.71

* $p < .05$.
** $p < .01$.
*** $p < .001$.

TABLE 4

MULTIPLE-REGRESSION ANALYSES

Accepted personal-history factor	Partial r	Simple r	Beta weight
Criterion: Paired-comparison performance rating (multiple R = .42)			
5. Financial Responsibility	.33	.33	.34
4. Leadership and Group Participation	.14	.14	.14
13. Vocational Satisfaction	.14	.06	.14
7. Parental Family Adjustment	−.12	−.08	−.11
6. Early Family Responsibility	.12	.18	.11
1. School Achievement	−.10	−.05	−.10
Criterion: Mean sales volume rank (multiple R = .50)			
5. Financial Responsibility	.31	.43	.33
8. Stability	.21	.39	.22
15. General Health	−.12	−.05	−.10
2. Higher Educational Achievement	.12	.12	.10
1. School Achievement	−.10	−.08	−.09
Criterion: Maximum sales volume rank (multiple R = .36)			
5. Financial Responsibility	.20	.31	.22
8. Stability	.12	.27	.13
7. Parental Family Adjustment	−.11	−.11	−.10
6. Early Family Responsibility	.10	.17	.10
Criterion: Route difficulty (multiple R = .27)			
3. Drive	.19	.16	.19
11. Educational-Vocational Consistency	.17	.12	.17
6. Early Family Responsibility	−.12	−.09	−.11
15. General Health	−.11	−.11	−.11
Criterion: Tenure as a salesman (multiple R = .30)			
13. Vocational Satisfaction	−.20	−.18	−.20
6. Early Family Responsibility	−.19	−.18	−.18
12. Vocational Decisiveness	.16	.12	.16

yses on all three criterion measures, the factor Financial Responsibility carried the greatest weight. For both the sales volume criterion measures, the next greatest weight was on the factor Stability. Interestingly enough, however, the second highest weighted factor for the performance-review criterion, which involved the judgment of the men's superiors, was Leadership and Group Participation. The factors Parental Family Adjustment (negative), Early Family Responsibility, and School Achievement each had weights in two of the three analyses. Again, it should be noted that all factors mentioned are among the first eight—those recommended for operational use.

The two remaining criterion measures—route difficulty and tenure as a salesman—yielded multiple correlations of .27 and .30, respectively. These coefficients are admittedly modest. However, they do indicate the flexibility of the background dimensions in predicting a number of different criteria, since the correlations of the two minor criteria with the three major ones are either zero or insignificantly negative.

In the route-difficulty analyses, the factor Drive carried the greatest weight. This is logical, since a high score on Drive, as Drive is defined here, would be associated with a willingness to take on the difficult and challenging sales routes. Enquiries to the organization did indeed elicit the fact that the older, less successful salesmen were generally to be found on the easier routes, which demanded mainly a servicing of established outlets rather than any real sales activity in opening new accounts.

The highest weight in the tenure analysis is a negative one on the factor Vocational Satisfaction. This apparently startling result can be explained on the basis of the factor definition and the particular sales force involved. In this rather old and established sales group (average tenure is over 10 yr.), the salesmen of longest standing are generally the men on the less difficult, servicing routes who no longer strive for promotion or for self-improvement through relocation. As a result, they are probably resigned to, rather than satisfied with, their choice of occupation and have no reason to feel that their period of peak performance is some time in the future.

Results of this nature emphasize the need for care in interpreting personal-history data

for purposes of predicting on-the-job behavior. The "blind" application of an "index" derived from the items in this factor would, it is true, probably result in a sales force with low turnover. However, that sales force would tend to consist of older indivduals who would feel they were past their peak performance, whose aspiration levels were low, and who were more interested in simply maintaining their present position than in advancing in the organization.

DISCUSSION

On the basis of these results, the picture that emerges of the successful salesman or sales manager is one of a man with a background of competent handling of his personal economy, an early vocational start with prime or sole responsibility for managing family finances, and, particularly for the managers, a past history of sales achievement and present stability in the work and family situations.

An interesting finding in the analysis of the criterion measures of on-the-job behavior was that the greatest weight for tenure was a negative one on the factor Vocational Satisfaction. This latter finding leads to the conclusion that it would be inexpedient to predict tenure on the grounds of vocational dissatisfaction unless the organization is prepared to maintain a stable, low-turnover sales force at the expense of having it composed of men who feel they cannot improve their performance either in other positions or in other companies. Further research should make it possible to define and measure objectively the different background dimensions which characterize successful performers in a variety of occupations.

A general aim of this study was to stress the need for a rational as well as an empirical approach to the use of personal background data. An attempt was made, therefore, to interpret the significance of the personal background factors in the life of the individual and to show their relationships to successful performance in different occupational settings. By contrast, the more frequently used empirical approach will yield, for each occupational group studied, a frequently highly discriminative but uninterpretable composite score or index.

REFERENCES

BAEHR, M. E., BURNS, R. K., & MCMURRY, R. N. *Personal history index*. Chicago: Industrial Relations Center, University of Chicago, 1965.

BAEHR, M. E., & WILLIAMS, G. B. Underlying dimensions of personal background data and their relationships to occupational classification. *Journal of Applied Psychology*, 1967, **51**, 481–490.

BUEL, W. D. Biographical data and the identification of creative research personnel. *Journal of Applied Psychology*, 1965, **49**, 318–321.

DUNNETTE, M. D. Personnel management. In P. R. Farnsworth (Ed.), *Annual review of psychology*. Vol. 13. Palo Alto: Annual Reviews, 1962. Pp. 285–313.

DUNNETTE, M. D., & MAETZOLD, J. Use of the weighted application blank in hiring seasonal employees. *Journal of Applied Psychology*, 1955, **39**, 308–310.

LOCKWOOD, H. C., & PARSONS, S. O. Relationships of personal history information to the performance of production supervisors. *Engineering Industrial Psychology*, 1960, **2**, 20–26.

MINOR, F. J. The prediction of turnover of clerical employees. *Personnel Psychology*, 1958, **11**, 393–402.

MORRISON, R. F., OWENS, W. A., GLENNON, J. R., & ALBRIGHT, L. E. Factored life history antecedents of industrial research performance. *Journal of Applied Psychology*, 1962, **46**, 281–284.

MOSEL, J. N., & WADE, R. R. A weighted application blank for reduction of turnover in department store sales clerks. *Personnel Psychology*, 1951, **4**, 177–184.

SCOLLAY, R. W. Validation of personality history items against a salary increase criterion. *Personnel Psychology*, 1956, **9**, 325–336.

SMITH, W. J., ALBRIGHT, L. E., & GLENNON, J. R. The prediction of research competence and creativity from personal history. *Journal of Applied Psychology*, 1961, **45**, 59–62.

(Early publication received August 11, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 2, 104–121

GOALS AND INTENTIONS AS MEDIATORS OF THE EFFECTS OF MONETARY INCENTIVES ON BEHAVIOR [1]

EDWIN A. LOCKE [2] AND JUDITH F. BRYAN

American Institutes for Research, Washington Office

AND LORNE M. KENDALL

Simon Fraser University, Burnaby, British Columbia

Based on the assumption that goals and intentions are the most immediate determinants of an individual's behavior, it was hypothesized that monetary incentives would affect task performance only through or by means of their effects on the individual's goals or intentions. 5 experiments were performed to explore this hypothesis. 2 dealt with the relationship of performance goals to level of performance (output) on a task as a function of incentive condition. 3 experiments examined the relationship of behavioral intentions to task choice as a function of incentive. In all 5 studies, significant relationships were obtained between performance goals or behavioral intentions and behavior. However, when goal or intention level was controlled, there was no effect of monetary incentive on behavior. In the 3 choice studies where incentives did have an initial effect on choice, these choice differences were accompanied by equivalent differences in intentions. When the latter were partialed out, the original incentive effect was vitiated. The data were interpreted as supporting the hypothesis.

A recent review of the literature on the motivational effects of financial compensation (Opsahl & Dunnette, 1966) stressed the need for more basic research on the mechanisms by which money motivates action. The authors stated that:

> . . . there is probably less solid research in this area than in any other field related to worker performance. We know amazingly little about how money either interacts with other factors or how it acts individually to affect job behavior. . . . Speculation, accompanied by compensation fads and fashions, abounds; research studies designed to answer fundamental questions about the role of money in human motivation are all too rare [p. 94].

The present studies were designed to help remedy this situation. The authors' approach was based on Ryan's (1958) theory of motivation [3] which states that: "a very large proportion of behavior is initiated by tasks [goals, intentions] and a very large proportion of tasks lead to the behavior specified by the tasks. . . . A task is a *necessary* condition for most kinds of behavior [p. 79]." If tasks, goals, or intentions are a necessary condition for action, then incentives such as money should affect action only if and to the extent that they affect the individual's goals and intentions. This was the major hypothesis tested in the studies to be reported here.

While the above represents little more than refined common sense, it should be recognized that the use of conscious goals and intentions as explanatory concepts in psychology is virtually nonexistent today. It is much more common to view incentives ("reinforcements") as shaping behavior automatically (without the operation of conscious volition). The present studies provided an opportunity to test the validity of the latter assumption.

Goals can be differentiated both qualitatively (e.g., "do your best" versus "try to

[1] This research was supported by contract Nonr 4792(00) between the Office of Naval Research and the American Institutes for Research. The views expressed are not necessarily those of the Navy Department. The authors would like to thank Charles D. Spielberger of Florida State University for his helpful criticisms of an earlier draft of this manuscript.

[2] Primary affiliation now with the University of Maryland.

[3] For a more recent statement of Ryan's theory, see also unpublished mimeos, 1964. Chapter I: Explaining behavior; Chapter II: Explanatory concepts; Chapter V: Experiments on intention, task, and set; Chapter VI: Intentional learning; Chapter VII: Unintentional learning. Cornell University, Department of Psychology.

improve") and quantitatively (e.g., "try for a score of 50" versus "try for a score of 70"). In the latter case, goal level would be synonymous with "level of aspiration." Two of the present investigators have reported a number of studies showing strong relationships between *S*'s goals and level of performance on laboratory tasks. These results indicate that hard goals produce a higher performance level (greater output) than easy goals (Bryan & Locke, 1967b; Locke, 1966a, 1966b; Locke & Bryan, 1966b, 1967a, 1967b); and that specific (quantitative) hard goals lead to a higher performance level than a goal of "do your best" (Bryan & Locke, 1967a; Locke, 1967; Locke & Bryan, 1966a, 1966b, 1967a, 1967b). A number of studies by other investigators also support the assumption that goals and intentions are important regulators of task performance (DeNike, 1965; Dey & Kaur, 1965; Dulany, 1962; Mace, 1935; Stedry, 1962).

Several recent studies of nonmonetary incentives have suggested that such incentives affect choice behavior only through their effects on the individual's conscious goals and intentions. For instance, Spielberger, Berger, and Howard (1963), Spielberger, Bernstein, and Ratliff (1966), and Spielberger, Levin, and Shepard (1962) found that the effect of *verbal reinforcements* on verbal behavior was a direct function of *S*'s awareness of the concept to be learned and the *desire* of *S* to get the reinforcement. Dulany (1962) and Holmes (1966) found that the individual's verbal responses on such tasks were a direct function of his behavioral intentions.

Locke and Bryan (1966b), using post-experimental goal questionnaires, found that knowledge of score had no effect on performance but that goals were significantly related to output. Locke (1967) noted that previous studies in this area often confounded goal-setting and knowledge effects. The *S*s given knowledge of score were typically given a specific hard goal to aim for, while *S*s not given knowledge of score were typically told to do their best. When these two factors were separated by independent experimental manipulation of goals and knowledge of score, Locke (1967) found that specific hard goals produced a higher level of performance than a goal of "do your best," but that there was no effect of knowledge of score per se on performance.

Goal setting as a mediator of the effects of *time limits* on work speed was investigated by Bryan and Locke (1967b). It was found that time taken to complete a task expanded to fill the time available (in line with Parkinson's Law) or contracted to get done in the time allowed, but only to the extent that *S*s set different goals under the different time limits.

Finally, Meyer, Kay, and French (1965) found that *participation* in performance appraisal interviews was relatively unlikely to lead to improved performance unless specific improvement goals were set during the interviews. The authors write:

> While subordinate participation in the goal-setting process had some effect on improved performance, a much more powerful influence was whether goals were set at all. . . .
>
> The average percent accomplishment estimate for those performance items that *did* get translated into goals was 65, while the percent estimate for those items that *did not* get translated into goals was about 27 [pp. 126–127]!

To the authors' knowledge, no studies have been made of goals and intentions as mediators of the effects of monetary incentives, although preliminary findings by Spielberger, Ratliff, and Bernstein (1966) suggested that the *desire* of *S*s to earn money on a verbal learning task played a role in determining its effects on performance. Whyte (1955) has shown in real-life settings that the effects of monetary incentives on output are by no means automatic. A series of behavioristically oriented experimental studies of monetary incentives by Toppen (1965a, 1965b, 1965c, 1966) found the following relationships between monetary reinforcement schedule and performance output on a lever-pulling task: (*a*) larger reward magnitudes and higher reinforcement frequencies led to higher output, (*b*) piece-rate payment yielded a higher output than time-rate payment, and (*c*) decreasing the magnitude of the reward over time resulted in performance decrement. Although Toppen asserts repeatedly in these studies that humans respond "just like animals" to reinforcement, his own findings and experi-

mental procedures consistently contradict this claim. Toppen repeatedly refers to the effects on performance of the *Ss*' "feelings of obligation" to *E*, their "expectations" and "interests," their "ulterior motives," and their "desire to do well," and he uses elaborate instructions and procedures to try to control and eliminate such effects. His very efforts in this respect reflect the importance and ubiquity of conscious factors in human performance.

What Toppen ignored was the fact that between the offering of a monetary incentive and an individual's overt action, a complex sequence of conscious and subconscious processes occur. The possibility that these processes might determine *how* an individual will react to the incentive was evidently not considered. For instance, before an external object (which does not physically touch *S*) can affect an individual's actions he must (ordinarily): *perceive* and identify it, that is, be aware *that* it exists and of *what* it is; following perception and identification comes an *appraisal* of the object against some standard of value, that is, an estimate of the extent to which the object will be beneficial or harmful to him. As a result of this appraisal (which may be entirely subconscious), he may experience an *emotion: "the psychosomatic form in which man experiences his estimate of the beneficial or harmful relationship of some aspect of reality to himself* [Branden, 1966, p. 5]." On the basis of his emotional reaction, his conscious values, and his perception of the task and situation, the individual will usually (explicitly or implicitly) set himself a *goal* or *end* toward which he will direct his behavior and in terms of which its appropriateness will be judged. He will develop an *intention* to act in a certain way.

It is assumed that the individual's specific goals and intentions are the end result or culmination of the complex intervening mental processes described above and the most direct determinants of actual performance. The present studies were not concerned with the question of *why* particular goals or intentions were developed by particular individuals, but only with the relationship between such goals or intentions, once established, and performance. To use Ryan's (1958) terminology, the authors were concerned with specific or immediate level explanation.

To demonstrate that goals and intentions were direct determinants of performance and mediators of the effects of incentives on behavior, it would have to be shown: (*a*) that goals and intentions were related to behavior regardless of incentive condition, but (*b*) that the reverse was not true—that incentives did not affect behavior unless they affected goals and intentions. The above requirements can be expressed more precisely in terms of three specific subhypotheses which were the ones tested in the present studies:

Hypothesis 1. Goals and intentions will be related to behavior regardless of incentive condition; that is, goals and intentions will be related to behavior, both within and across different incentive conditions.

Hypothesis 2. When incentive differences *do* correlate with behavior differences, these differences will be accompanied by corresponding differences in goals or intentions.

Hypothesis 3. When goal or intention differences are controlled or partialed out, there will be no relationship between incentive condition and behavior.

Hypothesis 1 is simply a restatement of the general assumption that goals and intentions are direct determinants of behavior. If this assumption is valid, then the relationship should hold up regardless of incentive condition. Hypotheses 2 and 3 are directly concerned with the role of goals and intentions as mediators of the effects of incentives. If incentives do affect behavior, they should affect goals and intentions at the same time (Hypothesis 2); if goal or intention differences are controlled, incentives should have no effect (Hypothesis 3).

The above three hypotheses were tested with respect to two different dimensions of behavior: level of performance or quantity of output, and direction of behavior or choice. Two experiments were performed in which all *Ss* worked on the same tasks under different incentive conditions, and *level of performance* (productivity) was the dependent variable. The relevant motivational concept in this situation is the individual's *performance goal* or *aspiration level.* If monetary incentives affect performance automatically then different

amounts of incentive should produce different levels of performance regardless of the level of the *Ss*' goals (e.g., the higher the incentive, the higher the performance level). On the other hand, if conscious goals mediate the effects of incentives, then incentive differences should only produce performance-level differences if and to the extent that they produce goal-level differences.

Three experiments were performed in which *Ss* had a choice as to the difficulty of the tasks they would work on, and *choice difficulty* was the dependent variable. Incentives were offered for solving a task (problem) successfully regardless of the difficulty of the task. This meant that *Ss* were "reinforced" with money more frequently for choosing easy tasks than for choosing hard tasks.

The relevant motivational concept in a choice situation is the individual's *behavioral intention* (Dulany, 1962), the choice or response that the individual intends to make. If monetary incentives affect task choice automatically, then *Ss* who are paid for success only should be more likely to choose easy tasks as the payment for success increases, regardless of their intentions. On the other hand, if behavioral intentions mediate the effects of incentives on task choice, then incentive differences should produce task-choice differences only if and to the extent that they produce intention differences.

The first two experiments deal with level of performance and the last three with choice. The first, fourth, and fifth experiments were designed and carried out by the first author; the second was designed and carried out by the second author; the third study was the responsibility of the third author.

EXPERIMENT 1

Method

Subjects

The *Ss* were 127 paid, undergraduate volunteers (71 males and 56 females) recruited from the University of Maryland. All *Ss* received $3 for participation in addition to whatever money (if any) they earned in the experiment.

Task

The task was giving uses for objects and was introduced as a test of creativity. On each trial, *S* was given a common object (e.g., a cardboard box), the task being to list possible uses for the object. There were two practice trials followed by three blocks of seven experimental trials each. Trials lasted 1 min. The *Ss* wrote their responses on answer sheets. During the experiment, *Ss* kept track of their own scores with the instruction that the protocols would be rescored after the experiment. The *E* did rescore them, eliminating only those responses that were not uses (e.g., if *S* said "to sell" an object without telling why he sold it, it would not be considered a use). Thus, *S*'s score was the total number of uses given without regard to quality.

Procedure and Conditions

After the rules were explained to *Ss*, they were given the two practice trials and told to do their best. On the first block of seven experimental trials, all *Ss* in Conditions A through E (see below) were told to "work at what you would consider to be a normal pace for an eight hour day, assuming no time limits and that you would not be paid for production." This block was of no direct relevance to the present study and will not be discussed further. The experimental conditions were introduced before Blocks II and III. The *Ss* in each condition were run together.

Condition A: Assigned goals; all-or-none incentive ($N = 19$). Before Block II, *Ss* were assigned the goal of attaining, on each trial in that block, a score equal to one more than their mean score on the two practice trials. They were offered a 5¢ bonus for each trial on which they reached this goal.

Before Block III, *Ss* were assigned a goal of attaining, on each trial, a score equal to one more than their best score in Block II. They were offered a 10¢ bonus for each time they reached or exceeded this goal.

Condition B: Assigned goals; no incentive ($N = 17$). This condition was identical to Condition A, except *Ss* were not offered any bonus for goal attainment.

Condition C: Self-set goals; piece-rate incentive ($N = 22$). The incentives were set for this group so that if they performed at the same level as Group A, they would receive, on the average, the same total pay on each block as Group A had received. Thus *Ss* were offered 6/10 of a cent per "use" on each trial on Block II and 4/10 of a cent per "use" on Block III. So that their goal level could be determined, *Ss* were told to indicate what their goal would be before each block. They were told that the goal would have nothing to do with their pay, which depended upon the number of uses given, but that they should indicate the goal (score) they were actually going to try for on each of the seven trials in that block.

Condition D: No goal setting; piece-rate incentive ($N = 21$). This condition was identical to Condition C, except that *Ss* did not set any explicit goals each block.

Condition E: Self-set goals; no incentive ($N = 12$). This condition was identical to Condition C, except that no incentive for performance was offered.

Condition A′: Assigned goals; all-or-none incentive ($N = 21$). This condition was similar to Condition A above, except that the goals in this case were set higher than for Group A. On Block I the A′ condition *Ss* were told to do their best. Each *S*'s Block II goal, for each trial, was one more than his best score on Block I, and his Block III goal, for each trial, was one more than his best score on Block II. The incentives were 10¢ and 25¢ for reaching or exceeding the goals on Blocks II and III, respectively.

Condition B′: Assigned goals; no incentive ($N = 15$). This condition was identical to Condition A′ above, except that no incentives were offered for goal attainment. Thus this condition holds the same relationship to Condition A′ as Condition B holds to A above.

All *Ss* were administered goal-description questionnaires at the end of the experiment to determine whether they had been trying for their assigned or self-set goals and to what degree they had been trying for the incentives. The responses indicated that nearly all *Ss* were trying for their goals and were trying to maximize their earnings.

Results and Discussion

The data were initially analyzed using Groups A through E only. There were no significant performance differences among these five groups in raw means either on the practice trials (indicating the groups were of equal initial ability) or on Blocks II and III. However, the within-cells variance for these groups was inflated due to wide differences in ability of the *Ss* within each group. Thus subsequent analyses were done using only improvement scores: (*a*) from the practice trials to Block II, and (*b*) from Block II to Block III.

The performance means for each group for Blocks II and III are shown in Figure 1a and the mean goals for the same groups and blocks are shown in Figure 1b (no goals are shown for Group D, since they did not set explicit goals). In parentheses beside each point in Figure 1b are shown the percentage of times (trials) on which *Ss* in that group reached or beat their assigned or self-set goals on that block. It is evident that the higher the goals, the less likely *Ss* were to succeed in reaching them.

The improvement scores from the practice trials to Block II showed no differences among the five groups. Since there were no differences among these groups on the practice trials, and since the goal levels of all groups on Block II were similar (see Figure 1b), this finding is consistent with Hypothesis 3.

The improvement scores from Block II to Block III tell quite a different story, how-

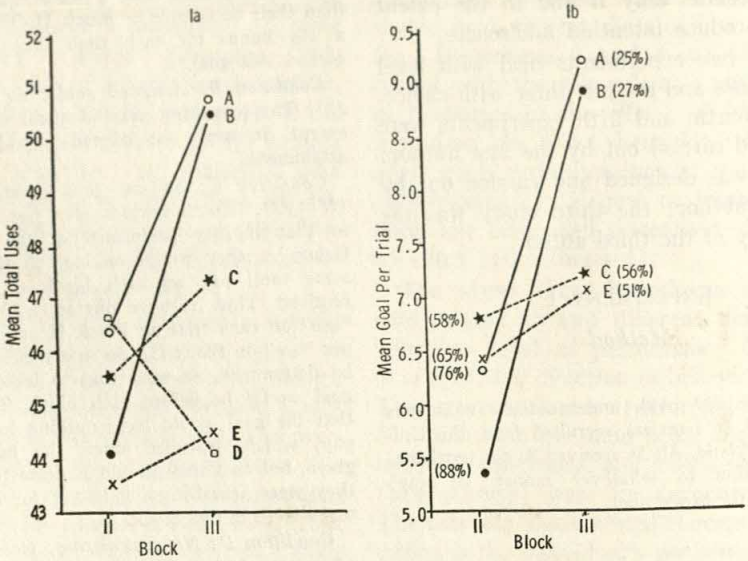


FIG. 1a and b. Performance means and mean goal-level scores for Groups A through E, by block (Experiment 1).

ever. If all five groups are included in the analysis, the F ratio for the difference in performance improvement means is 5.44 ($df = 4/86$), which is significant at the .001 level. It is clear from Figure 1b that the assigned goal groups (A and B) had considerably higher and harder goals on Block III than the self-set goal groups (C and E).

To determine the relative importance of goals and incentives in regulating performance, Groups A, B, C, and E were put into a 2×2 analysis-of-variance design, the two factors being assigned goals versus self-set goals (A and B versus C and E), and monetary incentive versus no monetary incentive (A and C versus B and E). The first analysis was done using *goal improvement* scores (from Block II to III) as cell entries to determine differences in goal level among these groups. Using the unequal N method described by Winer (1962, pp. 241ff), a significant goal effect was obtained ($F = 93.48$, $df = 1/66$; $p < .001$) in favor of the assigned goal group. There was no difference in goal improvement between the incentive and no-incentive groups and no interaction effect. If performance differences were confined to groups having significant goal-level differences, one would expect the only significant performance improvement difference to be for this same comparison. A similar 2×2 analysis of variance using *performance improvement* scores (Block II to III) as cell entries yielded a significant goal effect, the assigned goal groups (A and B) showing greater performance improvement than the self-set (C and E) goal groups ($F = 8.02$, $df = 1/66$; $p < .01$); there was no effect of incentive and no interaction effect.

Group D did not set explicit goals and their mean performance improvement was significantly lower than that of Groups C and E combined ($t = 2.82$, $p < .01$). Evidently the procedure of goal setting itself enhanced performance as compared to no goal setting, regardless of incentive condition.

There was no difference in the goal-level changes of Groups A and B (assigned goal with and without incentive) nor was there a difference in performance between these two groups. Similarly, the mean goal levels of Groups C and E (self-set goal with and without incentive) were almost identical and so were mean improvement scores. Thus there were no performance differences among groups offered different amounts of incentive providing their mean goal levels were the same.

Added credence is given to the belief that incentive differences will not produce performance differences if the goal levels are the same, by data from Groups A′ and B′. The performance data for these groups for Blocks II and III are shown in Figure 2a, and the goal level and success data are shown in Figure 2b. Again there was no significant

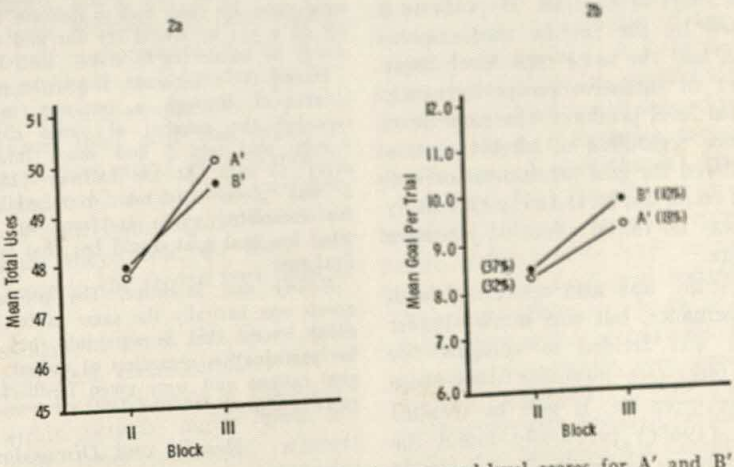


FIG. 2a and b. Performance means and mean goal-level scores for A′ and B′ groups, by block (Experiment 1).

difference between the two groups either in goal-level changes or in performance improvement from Blocks II to III. This replicates the similar findings for Group A versus B and Group C versus E, above.

An estimate of the overall between-group relationship of performance level to goal level was obtained by ranking the 12 performance means plotted in Figures 1a (omitting Group D) and 2a, and the 12 mean goal scores plotted in Figures 1b and 2b. The rank-order correlation between these pairs of means was .85 which is significant at the .01 level. The overall relationship between *change* in goal level and performance *improvement* from Blocks II to III was obtained using the 70 *S*s in Groups A, B, C, and E as the units of analysis. The cosine-pi estimate of r_{tet} (dichotomizing both variables at the median) was .55 which is significant at the .01 level. A similar relationship was not found among *S*s in Condition A′ and B′, however. This is probably due to the fact that the range of goal improvement scores was less among these *S*s. Improvement was harder for them since they were already at a high level of performance on Block II. As a result, the improvement scores may have been less reliable for these *S*s.

The results of this experiment supported Hypotheses 1 and 3. Hypothesis 1 was supported by the finding of a significant effect of goal level on performance level using both group means and individual change scores as the units of analysis. Hypothesis 3 was supported by the finding that, among groups which had the same goal level, there was no effect of incentive on performance. The same goal level produced the same level of performance regardless of whether incentives were offered for goal attainment or not. Hypothesis 2 could not be tested in this study since there was no initial effect of incentive on performance.

The next study was also concerned with level of performance, but was much simpler in design. It was decided to compare the effects of a piece-rate incentive with those of a time-rate incentive. It will be recalled that Toppen (1965c) previously found the former to produce a higher level of performance than the latter.

EXPERIMENT 2

Method

Subjects

The *S*s were 30 paid, male, undergraduate volunteers recruited from the University of Maryland. All *S*s were run individually. The *S*s were paid as indicated below.

Task

The task was assembling a number of identical objects from Tinker Toy sets. The object to be copied resembled a wind vane and was designed by the second author to insure a maximum utilization of all the Tinker Toy parts. The *S* was supplied with enough parts to make about 26 wind vanes, and *E* could take some apart if *S* made more than this in the time allowed. Boxes of like parts were placed in front of *S* to make assembly easier. The *S* had to make a complete wind vane before beginning another (no assembly lines were permitted).

Procedure and Conditions

The study was introduced as an experiment on the development of "manual dexterity" and "manipulative abilities." The *S* was shown the model and asked to make two wind vanes just like the model in as little time as possible. The model was placed in front of *S* and he could refer to it as needed. The *E* recorded the time taken to make each vane. At this point the experimental conditions were introduced.

Piece-rate incentive. Half the *S*s were told that they would get no basic (minimum) pay but would receive 12¢ for every vane they put together in the next 50 min. This rate was chosen to allow *S*s to make $3, on the average, in the 50-min. time limit. The *S* was given his fastest assembly time (of the two practice vanes) and told how many he could expect to make in 50 min. if he worked at that same pace. He then had to indicate to *E* the number of vanes he would try for and the amount of money he would try to make.

During the experiment, *E* left the room but could observe *S* through a one-way mirror. The *E* recorded the number of vanes completed every 5 min. and told *S* how much time had elapsed every 10 min. At the halfway (25 min.) point, *S* was given a 2-min. rest while *E* counted the completed vanes in front of *S* and asked what his final goal would be; that is, *S* set a new final goal.

Hourly rate incentive. The procedure for this group was basically the same as for the piece-rate group, except that *S*s were told they would get $3 for participation regardless of output. The *S*s made goal ratings and were given feedback as with the piece-rate group.

Results and Discussion

The two incentive groups did not differ significantly in initial ability as measured by

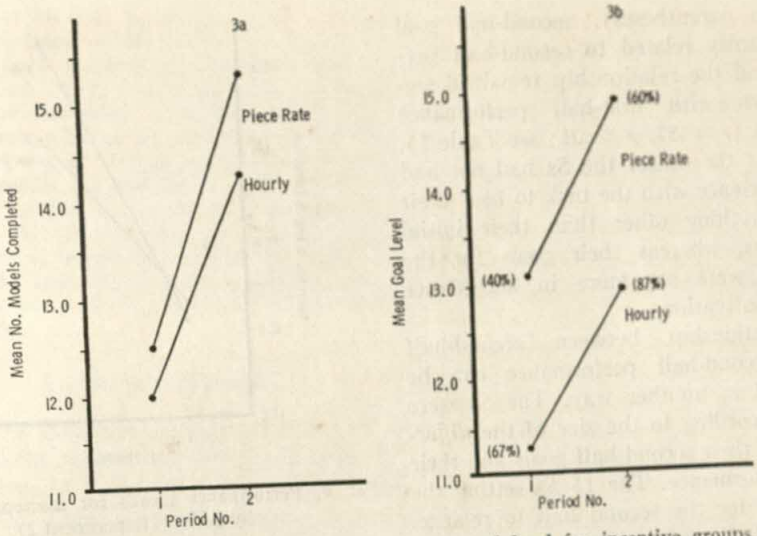


FIG. 3a and b. Performance means and mean goal level for incentive groups, by period (Experiment 2).

the speed with which they completed the fastest of their two practice models. Performance for the two groups during the first and second 25-min. periods of the experiment, measured in terms of the total number of models completed, is shown in Figure 3a. The goal levels of each incentive group for the first and second 25-min. periods are shown in Figure 3b. Since *S*s set their initial goals in terms of total output (performance for the whole 50-min. period), the first-period goals were obtained by dividing their initial total performance goals by two. Second-period goals were obtained by subtracting each *S*'s first-half performance score from his *revised total* output goal (made just before the second half of the experiment began). Since this revised total goal was made in full knowledge of first-period performance, it necessarily included *S*'s second-half performance goal. It is clear from Figures 3a and b that *S*s thought they could do more and actually did more in the second half than in the first half.

The percentages of times the *S*s in each incentive group beat their Period 1 and 2 goals are shown in parentheses in Figure 3b. Note that within periods, higher goals are associated with fewer successes, whereas across periods higher goals are associated with greater success. The latter finding is no doubt a result of learning: as *S*s became more proficient in the second period, they were able to set higher goals and to reach these goals more often as compared with the first period.

It appears from Figure 3a that the piece-rate group completed more models during each of the two periods than did the hourly group. However, *t* tests on these two pairs of means showed these differences to be well within the range of chance expectancy. Similarly, the piece-rate group, as shown in Figure 3b, appeared to set higher goals for themselves in both periods than did the hourly group, but neither of the differences exceeded chance level. A *t* test on the difference of the changes in productivity from the first to the second period showed no difference between the two incentive groups, nor was the difference in goal changes significant.

The *S*s in both incentive groups were combined in order to compute correlations between ability, first- and second-half performance, and first- and second-half goals. These correlations are shown in Table 1. All of the first-order correlations between the experimental variables were significant at the .01 level. Although first-half goal correlated significantly with first-half performance, when ability was partialed out the significant relationship disappeared (the partial correlation

is shown in parentheses). Second-half goal was significantly related to second-half performance, and the relationship remained significant even with first-half performance partialed out ($r = .57$, $p < .01$; see Table 1). Evidently, at the outset the *S*s had not had enough experience with the task to base their goals on anything other than their initial ability scores, whereas their goals for the second half were set more in accordance with their motivation.

This relationship between second-half goal and second-half performance can be demonstrated in another way. The *S*s were regrouped according to the size of the *difference* between their second-half goals and their first-half performance. The 15 *S*s setting the highest goals for the second half in relation to their first-half performance were compared with the 15 *S*s setting the lowest second-half goals in relation to their first-half performance. The difference in the mean discrepancy between the two selected groups was significant at the .01 level ($t = 6.92$), indicating that the division was successful. The first- and second-half performance of the two discrepancy groups is shown in Figure 4. The difference in performance improvement from the first period to the second period was significant at the .05 level ($t = 2.68$) in favor of the high-discrepancy group.

The results generally support Hypotheses 1 and 3. The first hypothesis was supported by the significant relationships found between second-half goal level and second-half performance for all *S*s combined and by the significant difference in performance improvement of the two discrepancy groups. Hypothesis 3 was supported by the finding that when the goal level of the different incentive groups was the same (not significantly different), the effect of incentive condition on performance was not significant. Once again, the incentives failed to produce any initial performance differences so that Hypothesis 2 could not be tested.

TABLE 1

CORRELATIONS BETWEEN GOALS AND PERFORMANCE FOR ALL SUBJECTS COMBINED

	Goal, first half	Goal, second half	Performance, first half	Performance, second half
Ability	.71[a]	.71	.79	.60
Goal, first half		.71	.63 (.16[b])	.62
Goal, second half			.93	.90 (.57[c])
Performance, first half				.85

Note.—Experiment 2; $N = 30$.
[a] All r's over .46 are significant at $p < .01$.
[b] With ability partialed out.
[c] With first-half performance partialed out.

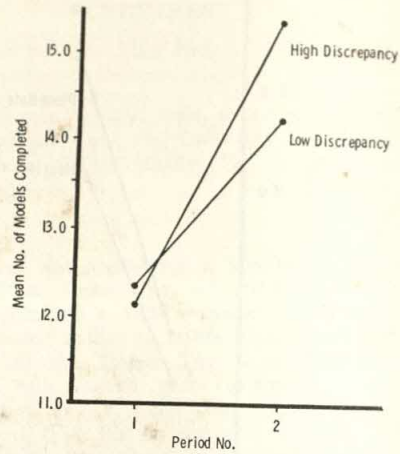


FIG. 4. Performance means for discrepancy groups, by period (Experiment 2).

EXPERIMENT 3

Method

Subjects

The *S*s were 22 Trenton State Teacher's College females who were members of an introductory psychology class. The *S*s volunteered for the experiment and were each paid $1 for participation at the end. However, they were told at the outset that they would get only what they earned during the experiment.

Task and Procedure

The task was word unscrambling. There were 6 practice trials followed by 3 blocks of 10 experimental trials each. On each trial, *S* was allowed to choose the *length* of word she would try to unscramble; *S* could choose any word length from 3 to 11 letters each time. The *S* had to make her choice before actually seeing the word but was told that words of the same length were equivalent in difficulty. After making her choice, *S* turned the page in her work booklet and worked on the word length she had chosen for 45 sec. On the first block of 10 trials, *S*s were not offered any incentive. On the second block, they were offered 2¢ for each word correctly solved regardless of the length of the word. On the third block of 10 trials, *S*s were

offered 10¢ for each word correctly solved regardless of the length of the word.

Before each *block* of 10 trials and before each *trial* of each block, *S* described her *behavioral intention* for that block or trial by indicating, on a 5-point scale, whether she would choose "the easiest word(s) possible," "fairly easy word(s)," "words that are neither too easy nor too hard," "fairly hard word(s)," or "the hardest word(s) possible" on that block or trial. The intention scale filled out before each *block* of words will be called the *B-intention* (for block) scale. The intention scale filled out before each *trial* will be called the *T-intention* (for trial) scale.

Results and Discussion

The *S*'s intention ratings were quantified on a 5-point scale (5 for the "hardest words" alternative, to 1 for the "easiest words" alternative). The dependent variable was the length of the word, in letters, chosen by *S* on a given trial or the mean word length chosen on a given block.

The mean within-*S* correlation (using Fisher *z* transformations for averaging) between T-intentions and word-length choice across the 30 trials was .81 (range = .12–.94; $Mdn = .80$). The mean within-block, between-*S*s correlation between B-intentions and mean word choice was .60 ($p < .01$). The corresponding mean within-block correlation of mean word choice with the mean T-intention was .77 ($p < .01$). B-intentions and mean T-intentions correlated .69 ($p < .01$), on the average, within blocks.

The mean word-length choices and intention ratings of the *S*s for each of the three blocks and the *t* ratios for change between blocks are shown in Table 2. Clearly, the higher incentives led to easier word choices. The proportions indicating the number of trials out of 10 on which *S*s succeeded in solving their chosen words were .60, .62, and .71 for Blocks I, II, and III, respectively. The changes in word-length choice were accompanied by equivalent changes in T-intention ratings. The B-intention ratings did not completely parallel those for word choice and mean T-intentions, but were in the same direction.

Across-block correlations between B-intentions, T-intentions, incentive, and word choice were obtained by treating the data from each block as though they were from different individuals (thus, the *N* for observations was 3 × 22 or 66). The incentive was coded 0, 1, or 2 (for the 0¢, 2¢, and 10¢ conditions, respectively). This gave maximum predictability to the incentive variable since the mean word-choice changes were about equal between Blocks I and II, and II and III, respectively (see Table 2). The across-block correlations between B-intentions, mean T-intentions, incentive, and mean word choice are shown in Table 3. Also shown are the intention–word-choice correlations with incentive partialed out and the incentive–word-choice correlations with intentions partialed out. For purposes of significance testing, an *N* of 22 was used in order to be conservative (in view of the fact that the 66 observations were not independent). The results indicated that both B-intentions and T-intentions were significantly correlated with word choice both before and after amount of incentive was partialed out. In contrast, while amount of incentive was related significantly to word choice before partialing, the relationship was vitiated when T-intentions were partialed out, or when both B-intentions and T-intentions were partialed out.

The results of this experiment support all

TABLE 2

MEANS AND *t* RATIOS FOR CHANGES IN INTENTIONS AND WORD-LENGTH CHOICE

Change	Means		
	Block I (0¢)	Block II (2¢)	Block III (10¢)
B-intentions	2.91	2.68	2.45
T-intentions	3.20	2.84	2.35
Word-length choice	6.32	5.68	5.08
	t ratios for mean change		
	Blocks I–II	Blocks II–III	Blocks I–III
B-intentions	1.31	1.05	2.35*
T-intentions	3.34**	3.34**	5.30**
Word-length choice	3.98**	3.03**	5.11**

Note.—Experiment 3; $N = 22$.
* $p < .05$.
** $p < .01$.

TABLE 3

RAW AND PARTIAL CORRELATIONS OF INTENTIONS AND INCENTIVE WITH WORD CHOICE

	T-intentions	Incentive	Word choice				
			Raw	Partial out			
				Incentive	B-intentions	T-intentions	B- and T-intentions
B-intentions	.69**	−.24*	.62**	.59**	—	—	—
T-intentions		−.50*	.84**	.79**	—	—	—
Incentive			−.51*	—	−.48*	−.21	−.22

Note.—Experiment 3, $N = 66_{obs}$; 22 *Ss*.
* Significance levels are based on $N = 22$. A negative correlation indicates the higher the incentive, the shorter the word lengths chosen or the "easier" the intention.
* $p < .05$.
** $p < .01$.

three of the hypotheses described earlier. (1) Intentions were related to choice within individuals, between individuals within-blocks (where incentive was constant), and between individuals across blocks (where incentive varied). (2) When incentive differences did produce choice differences, these differences were accompanied by equivalent differences in T-intentions (though not necessarily in B-intentions). (3) Finally, when across-block differences in intentions were partialed out, there was no relationship of amount of incentive to choice.

The next experiment was similar in design to this one except that only the B-intention scale was used and only two rather than three incentive conditions were included. In addition, on the 0¢ block in the next experiment, different instructions were given to different *Ss* and their effects on performance and intentions observed.

EXPERIMENT 4

Method

Subjects

The *Ss* were 70 University of Maryland, paid, undergraduate volunteers. Twenty-seven *Ss* were female and 43 were male. All *Ss* received $3 for participation plus whatever they earned in the experiment.

Task

The task was word unscrambling as in the previous experiment. In the present case, there were 3 practice trials followed by 2 blocks of 10 experimental trials each. Again *S* could choose *any* word length from 3 to 11 letters on each trial. The *Ss* chose their word length before actually seeing the word, but were told that all words at any given length were of equivalent difficulty.

Procedure and Conditions

Block I. The *Ss* were divided at random into three instructional groups on Block I. Twenty-three *Ss* were told to "try and achieve as much *success* as possible"; 23 *Ss* were told to "try and achieve as great a sense of personal accomplishment or personal *achievement* as possible"; and 24 *Ss* were told to "try and overcome the greatest possible *challenges* (even if you do not often make it)." No incentive was offered for success on this block.

Before the first experimental trial, *Ss* filled out 5-point *behavioral intention* scales, similar to those used previously, to indicate their intended word-choice difficulty for the next 10 trials (possible response categories ranged from "very easy words" to "very hard words").

After this, *Ss* worked on 10 different word choices for 45 sec. each, choosing whatever word length they wanted before each trial. The *Ss* were told that they were not committed to their intention ratings.

Block II. On Block II, all *Ss* were treated identically. All were offered 4¢ for each word correctly solved in this block regardless of the length of the word. Before this block, *Ss* filled out behavioral intention scales; they also indicated how much money they would try to earn on the coming block (level of aspired earnings). Again, *Ss* worked for 45 sec. on each word choice, selecting whatever word length they wanted before each trial.

Results and Discussion

The behavioral intention ratings were quantified on a 5-point scale (5 for the "very hard words" alternative, to 1 for the "very easy words" alternative). The Block I instructions were quantified on a 3-point scale

(5 for "challenge," 3 for "achievement," and 1 for "success"). The latter quantities were based on the authors' a priori notion of the difficulty of the choices these instructions would lead to if *Ss* followed them. The dependent variable was the mean word length chosen on a given block.

In Block I, instructions correlated .67 ($p < .01$) with mean word choice, while intentions correlated .88 ($p < .01$) with word choice. The correlation between intentions and word choice with instructions partialed out was .76 ($p < .01$); on the other hand, the correlation of instructions with word choice with intentions partialed out was only .08 (*ns*). Block II intentions correlated .81 ($p < .01$) with word choice on that block, while aspired earnings correlated .60 ($p < .01$) with the latter variable.

Figures 5a–c show the mean total word-length choices and mean intention ratings on each block, for the *Ss* in each Block I instruction condition ("success," "achievement," and "challenge"). Looking first at Figure 5a, the group trying for "success" on Block I chose easy words on that block (mean empirical probability of success = .76) and did not change their word choices substantially on Block II. The intention-level means for these *Ss* are also similar for Blocks I and II. Neither the word choice nor the intention means changed significantly between blocks.

Figure 5b shows corresponding data for the Block I "achievement" condition *Ss*. These *Ss* chose moderately difficult words on Block I (mean empirical probability of success = .48) and easier words on Block II. The t ratios for mean between-block changes were significant at the .001 level for both the word-length and intention-level measures (t's = 4.87 and 4.83, respectively).

The corresponding data for the Block I "challenge" *Ss* are shown in Figure 5c. These *Ss* chose relatively hard words on Block I (mean empirical probability of success = .33) and much easier words on Block II. The between-block mean changes in word-length choice and intention level were both significant at the .001 level (t's = 7.73 and 9.72, respectively).

The effect of controlling for intention level on the relationship of incentive to choice can be seen in Figure 6. Mean word-length choice is plotted as a function of incentive for each of three levels of intention (Intention Level 1 corresponds to the "very easy words" alternative; Level 2, to the "fairly easy words" alternative; Level 3, to the "words that are neither too easy nor too hard" alternative; not enough *Ss* checked the "fairly hard" and "very hard" words alternatives to make

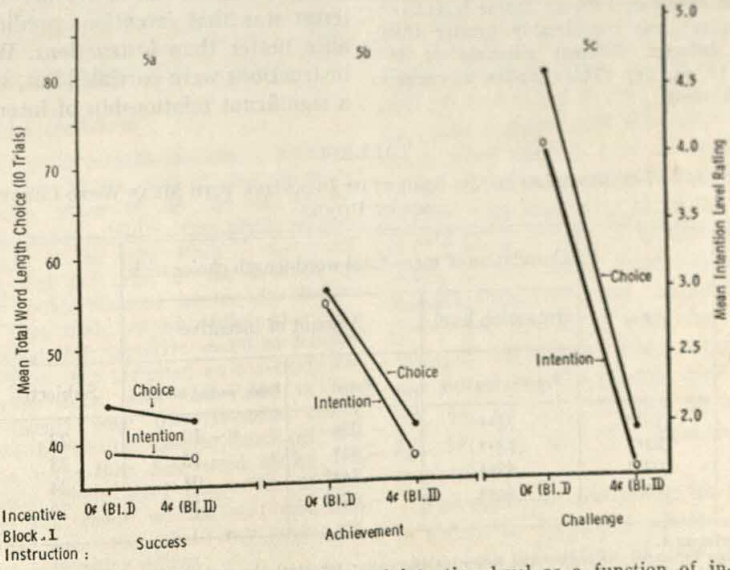


FIG. 5a–c. Mean word-length choice and intention level as a function of incentive and instruction condition (Experiment 4).

separate plots possible for them). The curves show a marked effect of intention level on word choice, but little effect of incentive on choice. None of the incentive effects is significant within any intention level.

The effects of controlling for intention can also be shown by means of partial correlation techniques. For this purpose, the intention, word-choice, and incentive scores from each block were treated as though they came from different individuals (making an N for observations of $2 \times$ the number of individuals involved). The incentive variable was coded 0 for Block I (where the incentive was 0¢) and 1 for Block II (where the incentive was 4¢). Thus, the correlations of incentive with the other variables were point-biserial coefficients.[4] The correlations of intention and incentive with word choice, and relevant partial correlations, are shown in Table 4 for each Block I instruction group and for all *S*s combined.

The results show that intention was significantly related to word choice both before and after incentive was partialed out, but that incentive was not related to word choice when intention was partialed out. This finding was true for each Block I subgroup as well as for all *S*s combined.

[4] It was believed that a biserial coefficient was not justified in this case, since the incentive classification did not represent coarse grouping but the real value of the two incentive conditions (0¢ and 4¢). In addition, the results of Experiment 5 indicated that the difference between "some incentive" and "no incentive" was considerably greater than the difference between different amounts of incentive (e.g., 1¢ vs. 2¢). This implies a genuine underlying dichotomy.

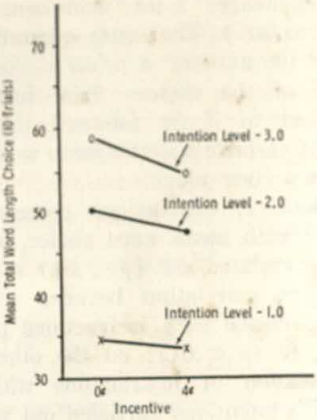


FIG. 6. Word choice as a function of amount of incentive, for three levels of intention (Experiment 4).

Like the previous experiment, these results support all three of our hypotheses: (1) intentions were significantly related to choice both within and across blocks; in addition, level of aspired earnings was correlated with choice on Block II where money was offered for success; (2) significant incentive effects on choice were accompanied by corresponding differences in intentions; (3) when the effect of intention level was controlled or partialed out, there was no relationship of amount of incentive to choice. A further finding of interest was that *intentions* predicted performance better than *instructions*. When Block I instructions were partialed out, there was still a significant relationship of intention to word

TABLE 4

CORRELATIONS OF INTENTION LEVEL AND AMOUNT OF INCENTIVE WITH MEAN WORD-LENGTH CHOICE, ACROSS BLOCKS

Subjects	r	Correlation of mean total word-length choice with			N	
		Intention level	Amount of incentive			
		$r_{\text{partial incentive}}$	r_{pbis}	$r_{pbis,\ \text{partial intention}}$	Subjects	Observations
Success	.78**	.78**	−.08[a]	−.10	23	46
Achievement	.81**	.74**	−.51*	−.13	23	46
Challenge	.93**	.82**	−.76**	.01	24	48
All subjects	.89**	.86**	−.48**	−.10	70	140

Note.—Experiment 4.

[a] Negative sign indicates the higher the incentive the easier (shorter) the word-length choice. Since there were only two values of incentive, 0¢ and 4¢, r's for incentive are point-biserial coefficients.

* $p < .05$ (based on N for *S*s).

** $p < .01$ (based on N for *S*s).

choice on that block, but there was no relationship of instructions to choice when intentions were partialed out.

While these results strongly replicated those of the first study of choice, both studies shared possible methodological flaws. First, the effects of order were confounded with the effects of amount of incentive, since the larger incentive always came later. Although this would not affect the within-block correlations, it could affect the mean changes in choice. Second, in both previous studies there were five intention categories, but only two or three levels of incentive. Thus, it might be argued that incentives did not get a "fair shake" because they had less chance to vary than did intentions. The final experiment on choice was designed to remedy these difficulties.

EXPERIMENT 5

Method

Subjects

The *Ss* were 64 University of Maryland, paid, undergraduate volunteers. Thirty *Ss* were female and 34 were male. The *Ss* received $3 for participation plus whatever they earned in the experiment (if anything).

Task

The task was word unscrambling as in the previous study. The *Ss* were given nine practice words (one at each length from 3 to 11 letters) followed by five blocks of five experimental trials (words) each. On each experimental trial, *S* could choose any word length from 3 to 11 letters.

Procedure and Conditions

Fifty *Ss* were in the *variable-incentive condition:* on each of the five blocks, each *S* was offered a different incentive, either 1¢, 2¢, 3¢, 4¢, or 5¢, for each word correctly solved on that block, regardless of the length of the word. Each *S* was in all five conditions, but the order of the 5 levels of incentive was counterbalanced across *Ss*. Before each block of five trials, an instruction sheet informed *Ss* of what their incentive would be for the coming block. They then indicated on a 5-point *behavioral intention* scale (like that used in the previous experiment) what their intended choice difficulty would be for the coming block of five trials. Next, *Ss* made their first word choice and worked on their chosen word length for 45 sec. before making their choice for the next trial and so on. The *Ss* were told that they were not committed to their intention ratings.

Fourteen *Ss* were in the *no-incentive condition.* These *Ss* were not offered any incentive on any block, but filled out an intention scale like that of the variable-incentive *Ss* before each block of five trials. At the end of the fifth block, all *Ss* were asked to describe, in their own words, what their overall intentions had been during the experiment.

Results and Discussion

The behavioral intention ratings were quantified on a 5-point scale as previously, and the criterion was again the mean word-length choice of each *S* on each block. The incentive conditions were quantified on a 5- or 6-point scale (depending upon whether or not the no-incentive *Ss* were included) according to the amount of incentive offered.

The within-incentive correlations between intentions and word-length choice were .78, .83, .79, .85, and .79 for the 1¢, 2¢, 3¢, 4¢, and 5¢ blocks, respectively ($p < .01$ for all r's). The corresponding figure for the no-incentive *Ss* was .59 ($p < .05$), using blocks as the unit of analysis, and .61 ($p < .05$), using individuals (sums across blocks) as the unit.

Figure 7 shows the relationship of amount of incentive to intentions and word choice. The dotted lines connecting the no-incentive-group curves with the variable-incentive-group curves indicate that the two groups of *Ss* were different. There was a slight but significant tendency within the variable-incentive condition for *Ss* to choose easier words when a larger incentive was offered ($F = 2.78$, $p < .05$; $df = 4/196$). A similar effect was found using intentions as the dependent variable ($F = 3.00$; $p < .05$; $df = 4/196$). There was also a significant difference between the mean word choices (for all trials combined) of the variable-incentive *Ss* and the no-incentive *Ss*, with the latter choosing harder words ($t = 4.13$, $p < .01$). A similar difference was found in the mean intention scores of the variable-incentive and no-incentive groups ($t = 3.33$, $p < .01$), with the latter having "harder" intentions. The proportions of success in solving the chosen word lengths were .52, .68, .67, .71, .79, and .76 for the 0¢, 1¢, 2¢, 3¢, 4¢, and 5¢ incentive conditions, respectively.

The effect of controlling for intention level is shown in Figure 8 where word choice is plotted as a function of incentive for each of three levels of intention. The intention levels

correspond to the three "easiest" alternatives on the intention scale, except that the top category includes all intentions of 3 or higher. There was a marked effect of intention level on word choice, but very little effect of incentive within any given intention level. There was a tendency for no-incentive *S*s with intentions of 3 or more to choose harder words than variable (1¢ to 5¢) incentive *S*s with the same intentions ($t = 3.15$, $p < .01$), but the t ratio may have been inflated due to the fact that the observations within each group were not independent. (Since the unit of analysis was the block, rather than the individual, many *S*s appeared more than once in the analysis.)

Correlations between amount of incentive, intentions, and word choice were obtained using independent observations by the following procedure: for each *S*, his intention, mean word choice, and corresponding incentive scores were chosen for just *one* of the five blocks that he worked on. The data were chosen so as to equalize the representation of each of the five incentive levels and so as to control for order effects. The same procedure was used to select data from the no-incentive condition, except that only order had to be controlled. Thus, the correlations between intentions, incentive, and word choice are based on only *one-fifth* of the total data, but no *S* was entered twice in the analyses as in the previous two experiments. The relevant raw and partial correlations are shown in Table 5, for the variable-incentive group separately and for the variable- and no-incentive groups combined.

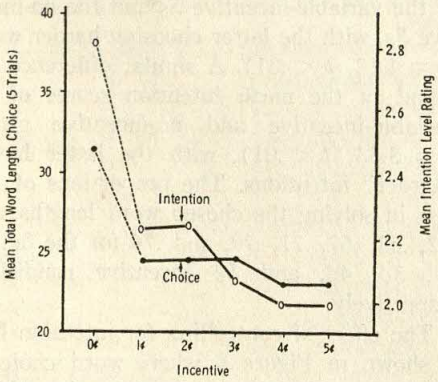


FIG. 7. Mean word choice and intention level as a function of incentive (Experiment 5).

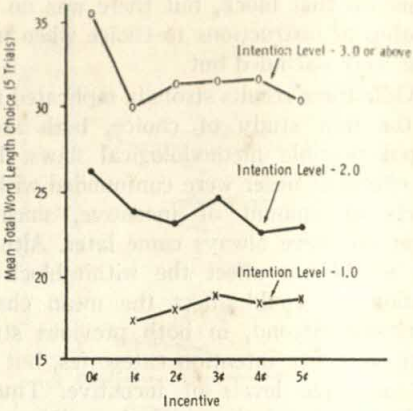


FIG. 8. Word choice as a function of amount of incentive, for three levels of intention (Experiment 5).

The correlations between intentions and word choice were highly significant ($p < .01$) both before and after partialing out incentive. In contrast there was very little relationship of amount of incentive to word choice before partialing out intentions, and no relationship after partialing intention level.

These findings clearly support those of the previous two studies, even with order effects controlled and using six different incentive levels: (*a*) there was a significant relationship between intentions and choice both within and between incentive conditions; (*b*) significant incentive effects on choice were accompanied by corresponding mean intention differences; and (*c*) when intention level was partialed out, there was little or no effect of incentive on choice.

A further finding of interest in this study emerged from the analysis of *S*s' postexperimental descriptions of their general intentions in the experiment. Two independent raters rated each intention description on a 1–5 scale according to the choice difficulty it suggested (e.g., an intention "to make as much money as possible" would get a score of 1, while an intention to "get a sense of accomplishment" would be given a score of 3). The correlation between the score assigned by the two raters was .90, for those 51 protocols to which both raters felt confident enough to assign scores. To obtain summary scores for these *S*s, the two scores were summed. Summary scores were obtained for an additional 8 *S*s, whose ratings were only assigned scores by one rater, by dou-

TABLE 5

CORRELATIONS OF INTENTION LEVEL AND AMOUNT OF INCENTIVE WITH MEAN WORD-LENGTH CHOICE, ACROSS BLOCKS

Subjects	Correlation of mean total word-length choice with				N
	Intention level		Amount of incentive		
	r	$r_{partial\ incentive}$	r	$r_{partial\ intention}$	
Subjects in 1–5¢ condition only	.83**	.83**	−.01*	.04	50
All subjects (1–5¢ and no-incentive condition combined)	.75**	.73**	−.29*	−.11	64

Note.—Experiment 5.
* Negative sign indicates the higher the incentive the easier (shorter) the word-length choice.
* $p < .05$.
** $p < .01$.

bling the latter score. For these 59 *Ss*, summary postperformance intentions correlated .68 ($p < .01$) with summary preperformance intentions (the latter were obtained by summing the five intention ratings *Ss* made during the experiment). Summary preperformance and postperformance intentions correlated .83 ($p < .01$) and .68 ($p < .01$), respectively, with mean word-length choice on the five blocks combined.

The former correlation was significantly greater than the latter (t for nonindependent r's $= 2.60$, $p < .05$), indicating that the postperformance measures were poorer "predictors" of actual behavior than preperformance measures. This would seem to contradict the claim frequently made by behavioristically oriented investigators that correlations between postperformance measures and performance are spurious because the former are only *descriptions* (or rationalizations) of behavior. If this were true, the postperformance measures should be at least *as* highly related to behavior as preperformance measures, if not more so.

CONCLUSION

The results of the five studies reported here are consistent with the hypothesis that goals and intentions are important determinants of behavior, and with the hypothesis that they are the mechanism by which monetary incentives influence behavior. All five studies showed significant relationships between goals or intentions and behavior within and/or between different monetary incentive conditions. Further, in each of the five studies it was demonstrated that if goal or intention level was controlled or partialed out, there was no effect of amount of incentive on behavior. This was demonstrated in the studies of performance level by showing that the same goal level produced the same performance level regardless of whether incentives were offered for performance or not. The same fact was demonstrated in the studies of choice by showing that when behavioral intention level was partialed out, the initial effects of incentives on choice were vitiated. Finally, the three studies of choice demonstrated that when incentives did affect behavior, these behavior differences were associated with equivalent differences in behavioral intentions.

The three studies of choice yielded a more complete test of the theory than did the two performance-level studies since the dependence of incentive effects on intention differences could be demonstrated directly. However, the fact that monetary incentives did not show any initial effect in the performance-level studies suggests the lesser potency of these incentives as compared with performance goals.

The first study suggests the possible importance of the manner in which incentives are arranged. All-or-none incentives made contingent upon reaching a given performance level were more effective, in this case, than the same payment rate made on a piece-rate

system. The reason seems clear: the goals *Ss* set themselves under the piece-rate system were lower than those that were assigned under the all-or-none system. There is no reason to believe that this will inevitably be the case, but it does point up the potential importance of the manner in which incentive systems are structured. Essentially, a piece-rate system lets the individual set his own overall work and pay goals, while an all-or-none payment system requires that he reach a specific level of performance before he gets anything.

Interestingly, back in 1911, Taylor implicitly recognized the potential difference between the two types of incentive system as well as the relationship of pay to goal setting. Under what Taylor called the traditional system of management by "initiative and incentive" the worker set his own pace, chose his own goal, determined his own work methods, and was paid accordingly, typically on a piece-rate system. Taylor's scientific management system introduced, among other things, performance standards or *specific work goals* for workers. Payment of the bonus (or the higher rates of pay which accompanied the introduction of his system) was contingent upon reaching assigned work goals. This goal was inevitably higher than the level of performance which had previously existed. Employees were told that they were expected to earn so much money (which meant produce so much) if they wished to remain on the job. Taylor called these "high-priced men."

Opsahl and Dunnette (1966) recognize the possible importance of goal setting in arguing that it might be beneficial to integrate pay-incentive systems with goal setting in industry, for example, by having pay increases dependent upon fulfillment of work goals (p. 104), or by having individuals actually set pay goals.

It may be hypothesized that assigned all-or-none goals or goals formally agreed upon by management and the worker will be more effective than self-set goals only to the extent that: (*a*) the former goals are set at a *higher level* than the latter, and (*b*) the former procedure results in a greater or *more intense commitment* of the individual to the goal.

This brings up the problem of the specific psychological role that monetary incentives play in motivation. Although such incentives appear not to influence performance directly, they may serve a number of other important functions:

1. Incentives can influence the *nature* of the intentions one develops on a task. If the payoff for succeeding at a hard task is large, the individual may be more likely to set a hard goal and take a big risk than if the payoff is small; or if the payoff is greater for succeeding at easier tasks (as in Experiments 3–5), this may lead him to establish easier intentions and to choose easier tasks.

2. Monetary incentives may persuade an individual to *accept* goals or tasks that are assigned to him. In Experiment 1, the "demand characteristics" (Orne, 1962) of the experimental situation were sufficient to persuade the assigned-goal, no-incentive *Ss* to accept their assigned goals, but such cooperation could not be taken for granted in most industrial situations. Workers will generally not accept a work goal or task unless it achieves other valued goals, such as to make money.

3. Monetary incentives may affect the *degree of commitment* of an individual to his goal or behavioral intention. Commitment may be expected to influence the degree of persistence an individual will show in the face of difficulty and frustration, the degree to which he will retain the goal if it conflicts with other goals (e.g., to be liked by coworkers), and the probability of his abandoning the goal altogether and "leaving the field" in the face of alternatives.

It must be stressed that whatever the effects of monetary incentives on performance, their *ultimate* impact should be a function of the degree to which the individual *values* money as compared to other incentives and his perception of the degree to which a given course of action is seen as a *means* of attaining this value (i.e., the perceived *instrumentality* of behavior in gaining the value; Vroom, 1964).

While the present research did not explore the evaluation process itself, the findings did indicate that attention should be paid to the goals and intentions which result from it. Many incentive programs are

based on the implicit assumption that incentives will affect behavior more or less automatically. The present findings show such an assumption to be vastly oversimplified and highly misleading, even in the context of laboratory experiments in which there is a high degree of control over the situational variables.

REFERENCES

BRANDEN, N. Emotions and values. *The Objectivist*, 1966, **5**(5), 1–9.

BRYAN, J. F., & LOCKE, E. A. Goal setting as a means of increasing motivation, *Journal of Applied Psychology*, 1967, **51**, 274–277.

BRYAN, J. F., & LOCKE, E. A. Parkinson's law as a goal-setting phenomenon. *Organizational Behavior and Human Performance*, 1967, **2**, 258–275. (b)

DENIKE, L. D. Recall of reinforcement and conative activity in verbal conditioning. *Psychological Reports*, 1965, **16**, 345–346.

DEY, M. K., & KAUR, G. Facilitation of performance by experimentally induced ego motivation. *Journal of General Psychology*, 1965, **73**, 237–247.

DULANY, D. E., JR. The place of hypotheses and intentions: An analysis of verbal control in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness*. Durham, N. C.: Duke University Press, 1962. Pp. 102–129.

HOLMES, D. S. Verbal conditioning or problem solving and cooperation? Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1966.

LOCKE, E. A. A closer look at level of aspiration as a training procedure: A reanalysis of Fryer's data. *Journal of Applied Psychology*, 1966, **50**, 417–420. (a)

LOCKE, E. A. The relationship of intentions to level of performance. *Journal of Applied Psychology*, 1966, **50**, 60–66. (b)

LOCKE, E. A. Motivational effects of knowledge of results: Knowledge or goal setting? *Journal of Applied Psychology*, 1967, **72**, 324–329.

LOCKE, E. A., & BRYAN, J. F. Cognitive aspects of psychomotor performance: The effects of performance goals on level of performance. *Journal of Applied Psychology*, 1966, **50**, 286–291. (a)

LOCKE, E. A., & BRYAN, J. F. The effects of goal-setting, rule-learning and knowledge of score on performance. *American Journal of Psychology*, 1966, **79**, 451–457. (b)

LOCKE, E. A., & BRYAN, J. F. Goals and intentions as determinants of performance level, task choice and attitudes. American Institutes for Research, Silver Spring, Maryland, AIR-E-51-1/67-FR, January 1967. (a)

LOCKE, E. A., & BRYAN, J. F. Performance goals as determinants of level of performance and boredom. *Journal of Applied Psychology*, 1967, **51**, 120–130. (b)

MACE, C. A. Incentives: Some experimental studies. Industrial Health Research Board (Great Britain), 1935, Report No. 72.

MEYER, H. H., KAY, E., & FRENCH, J. R. P., JR. Split roles in performance appraisal. *Harvard Business Review*, 1965, **43**, 123–129.

OPSAHL, R. L., & DUNNETTE, M. D. The role of financial compensation in industrial motivation. *Psychological Bulletin*, 1966, **66**, 94–118.

ORNE, M. T. On the social psychology of the psychological experiment with particular reference to demand characteristics. *American Psychologist*, 1962, **17**, 776–783.

RYAN, T. A. Drives, tasks, and the initiation of behavior. *American Journal of Psychology*, 1958, **71**, 74–93.

SPIELBERGER, C. D., BERGER, A., & HOWARD, K. Conditioning of verbal behavior as a function of awareness, need for social approval, and motivation to receive reinforcement. *Journal of Abnormal and Social Psychology*, 1963, **67**, 241–246.

SPIELBERGER, C. D., BERNSTEIN, I. H., & RATLIFF, R. G. Information and incentive value of the reinforcing stimulus in verbal conditioning. *Journal of Experimental Psychology*, 1966, **71**, 26–31.

SPIELBERGER, C. D., LEVIN, S. M., & SHEPARD, M. The effects of awareness and attitude toward the reinforcement on the operant conditioning of verbal behavior. *Journal of Personality*, 1962, **30**, 106–121.

SPIELBERGER, C. D., RATLIFF, R. G., & BERNSTEIN, I. H. Verbal conditioning with verbal and monetary reinforcers. *Psychological Reports*, 1966, **19**, 275–283.

STEDRY, A. C. Aspiration levels, attitudes, and performance in a goal-oriented situation. *Industrial Management Review*, 1962, **3**, 60–76.

TAYLOR, F. W. *The principles of scientific management*. New York: Harper, 1911.

TOPPEN, J. T. Effects of size and frequency of money reinforcement on human operant (work) behavior. *Perceptual and Motor Skills*, 1965, **20**, 259–269. (a)

TOPPEN, J. T. Money reinforcement and human operant (work) behavior: II. Within-subject comparisons. *Perceptual and Motor Skills*, 1965, **20**, 1193–1199. (b)

TOPPEN, J. T. Money reinforcement and human operant (work) behavior: III. Piecework-payment and time-payment comparisons. *Perceptual and Motor Skills*, 1965, **21**, 907–913. (c)

TOPPEN, J. T. Money reinforcement and human operant (work) behavior: IV. Temporally extended within-*S* comparisons. *Perceptual and Motor Skills*, 1966, **22**, 575–581.

VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.

WHYTE, W. F. *Money and motivation*. New York: Wiley, 1955.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received March 13, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 2, 122–126

EFFECTS OF CHANGES IN JOB-SATISFACTION LEVELS ON EMPLOYEE TURNOVER

CHARLES L. HULIN [1]

University of Illinois

The results of a program designed to increase the job satisfaction and decrease the turnover rate among a large group of female clerical workers are presented. Results indicated that levels of satisfaction were increased, with the largest increases occurring in the satisfaction variables stressed in the program. A significant decrease in turnover (from 30% to 12%) was also observed. Several alternative explanations for increased satisfaction and decreased turnover were considered, due to the lack of a completely adequate control group. None was considered adequate to explain the particular pattern of results obtained.

Previously reported data (Hulin, 1966) indicated the presence of an abnormally high turnover rate among the female clerical workers employed by the corporate offices of a large Canadian manufacturing firm. The turnover rate among the female employees of this company (30%/yr) was not only high on an absolute basis but was high relative to 15 other large firms located in Montreal who had a mean annual turnover rate of only 20%. The company involved ranked either first or second in turnover rate during each of the last 10 yr. among this group of companies. Many of the ecological factors which contribute to high turnover rates (general economic conditions of the area, labor market conditions, competition from other firms) and the characteristics of females in the work force (age, marital status, previous experience, domestic responsibilities, pregnancy, marriage) were uncontrolled but were common to the other firms in the area who had much lower turnover rates. Therefore, consideration was given to the role job satisfaction and dissatisfaction played in contributing to the high turnover rate.

The relationship between job satisfaction and turnover was studied (Hulin, 1966). Job-satisfaction questionnaires were administered to 345 female clerical workers employed at this company. The questionnaire measured satisfaction with six aspects of work. A year after the questionnaires were administered, 43 of the girls had quit who had completed the questionnaires and who could be identified on the basis of birth date, department, language, and marital status. A number of girls, who quit during the year following the survey, had started work after the survey in July 1964, a number did not report sufficient biographical information on the questionnaire so positive identification was possible, and some had changed departments.

For each of the 43 terminators, 2 control *S*s were drawn from the sample of girls who had remained on the job. These controls were selected so they matched the terminators on age, years of education, and mother tongue. A comparison of the mean satisfaction scores of these two groups indicated the two groups were significantly different in terms of group centroids. Correlational analyses indicated a multiple correlation between the satisfaction variables and turnover of .34 ($p < .01$) and a significant Pearson correlation between an unweighted sum of the satisfaction variables and turnover of $-.27$ ($p < .01$).

On the basis of these results the question was raised as to whether the turnover rate for this company could be reduced by increasing the job satisfaction of the clerical workers. In order to answer this question a program was initiated by the company to attempt to increase job satisfaction in the hope of reducing turnover. It is not the purpose of this article to analyze the program per se. However, a brief description may be in order. The results of the scores on the Job Description Index (JDI) indicated that dissatisfaction of the clerical workers was occurring in

[1] The writer would like to thank James Anderson and Susan Pfeifer for their assistance on the data-analysis stages of this research and Milton Blood who read and commented on an earlier draft of the paper.

all areas. (The JDI is a cumulative-point, adjective checklist type of scale. It has undergone extensive validation research and has been described elsewhere—Quinn and Kahn, 1967, p. 456; Vroom, 1964, p. 100. The JDI was constructed to measure five separate aspects of job satisfaction: satisfaction with work done, with the pay, with the supervisor, with promotional opportunities and policies, and with co-workers.) Analysis of their responses to open-ended questions substantiated this finding and also indicated that their dissatisfaction with pay was directed as much at the wage and salary *administration* as it was at wage *levels*. Examination of the workers' responses on the satisfaction with promotions scale indicated that the girls felt dissatisfaction was due to their feelings of being on "dead-end jobs" with little or no chance for promotion and not due to their feelings about the fairness of the policies. These analyses led the company to revise their wage and salary administration policies to make them more nearly constant across departments and to institute regular salary reviews and formalize the merit-raise procedure. Also a policy was initiated of allowing and encouraging intracompany transfers so the girls could move from one department to another where they would have a better chance of being promoted. In conjunction with this latter procedure the girls were encouraged to make changes in their own job so that they could have a more responsible job within their present job classification. The most compelling reason for choosing these job aspects to concentrate on in the program was because they are the job characteristics which are most fully under the control of the company. The entire program was put into effect in October 1964.

The purpose of the present article is to analyze the changes which have occurred in this company since July 1964 regarding satisfaction scores, turnover rates, and the relationship between satisfaction and termination decisions.

METHOD

Research Setting

The company in which this study was carried out has been previously described (Hulin, 1966; Hulin & Smith, 1967). Briefly, it is a large international manufacturing company which employs approximately 400 female clerical workers in the corporate offices. During the 3 yr. (1961, 1962, 1963) preceding initial study the turnover rate among the female clerical staff was 30.3%, 30.0%, and 30.0%. This company's rate of turnover was stable and had shown no tendencies in either direction over the past 10 yr. Turnover rate was computed by obtaining the ratio of number of quits per year to average number of female employees where average number of employees was equal to the average of the number at the start of the year and the number at the end of the year.

Subjects

The entire female clerical staff was asked to participate in the job-satisfaction survey which was to be conducted by the company. The workers were informed of this by their supervisors and by a letter from the personnel department. They were told that the questionnaires were completely anonymous and that their individual responses would never be revealed to the company. The questionnaires were given to the staff members by departmental representatives after the representatives had met with the investigator and had the questionnaires explained to them. The questionnaires were sealed in envelopes and returned either to the investigator at the offices or sent through the mail directly to the University of Illinois. Of the 350 members of the clerical staff present during the week of the survey, 298 (85.1%) participated in the survey. (Not all of the staff was present due to sickness, vacation, travel, etc.)

Variables

The job satisfaction of these clerical workers was measured by the Job Description Index (JDI). In addition to the five satisfaction variables, measures were also obtained of each worker's age, education level, job level (obtained by matching reported job title to the job-evaluation scale of the company), socioeconomic background (measured by placing the occupation of the father on an 11-point scale from professional to unskilled worker), and mother tongue. Measures of all variables were obtained during the first week of July 1966. By the first week of April 1967, 32 girls had quit, of whom 20 could be positively identified. However, the data of 4 of these 20 had to be eliminated from subsequent analysis because 3 of them were summer students and 1 of them did not report enough biographical information to yield measures of the control variables so that it was possible to obtain control *Ss*.

RESULTS

The turnover rate among female clerical workers during 1966 was 12%. This is significantly lower than the turnover rate of

TABLE 1

MEAN 1964 AND 1966 JDI SATISFACTION SCORES

Variable	Mean 1964 scores (n = 345)	1964 percentile equivalents	Mean 1966 scores (n = 298)	1966 percentile equivalents	t
Satisfaction with:					
Work	35.33	44	36.11	47	.85
Pay	15.01	14	32.83	66	10.27**
Promotions	10.78	48	24.58	74	10.38**
Co-workers	41.53	39	43.49	47	2.45*
Supervision	40.85	48	43.23	60	2.76**

* $p < .05$.
** $p < .01$.

30% in 1964 before the changes in the company's programs and policies were initiated. The turnover rate during 1965 was 18% indicating that there has been a steady decrease.

As in the 1964 survey, mean satisfaction scores were computed for each of the five job areas. Table 1 shows these mean satisfaction scores as well as those obtained in 1964. t tests of the significance of the difference between the 1964 and 1966 means were computed. The results of these t tests are shown in Table 1. Of the five job areas, four—pay, promotion, supervision, and co-workers—showed a significant increase in satisfaction from 1964 to 1966. Promotion and pay satisfaction showed the biggest increases. For both these variables the increase was significant beyond the .001 level. The mean satisfaction with work did not increase significantly from 1964 to 1966. While this analysis (multiple t tests done on correlated dependent variables) is not exactly correct and yields an underestimate of the α level for the results of the study, it is the most nearly correct analysis considering the hypothesis of the study. The concern was not with changes in the total set of dependent variables but with *where* the changes occurred. Further, a T^2 analysis, the multivariate analogue of the t test (Cooley & Lohnes, 1962), would not be appropriate either since the variance-covariance matrices from the 1964 and 1966 samples were not equal. This, of course, violates one of the assumptions of the T^2 analysis.

The columns headed "1964 and 1966 percentile equivalents" present the percentile equivalents of mean scores based, of course, on the appropriate numbers of *S*s. These percentile scores were computed with reference to norms which have been constructed on the basis of approximately 650 female workers from the United States work force.

For each of the 16 terminators who could be identified, 2 control *S*s were drawn who were matched in terms of age, years of education, job level, and mother tongue. A linear multiple-regression analysis was then performed on this group of 16 terminators and 32 control *S*s. A linear multiple correlation of .45 ($p < .05$) was obtained. An unweighted sum of the five satisfaction scores was correlated $-.46$ (r_{bis}) ($p < .01$) with the turnover.

The multiple correlation of .45 and the r_{bis} of $-.46$ predicting turnover indicate that even with turnover rates as low as 12% these termination decisions bear a relationship to the satisfaction of the workers. Further, the multiple R of .45 is not significantly different than the multiple R of .34 which was obtained in the 1964 analysis (Hulin, 1966). This indicates that a turnover rate of 12% in this setting does not represent the base rate which can be expected due to non-job-related variables.

DISCUSSION

The results of this study indicate that there was a decrease in turnover of 18% after the changes in the company's pay and promotion policies were instituted. In addition, there was a significant increase in satisfaction with four out of five of the job areas. Considering

these results, one is tempted to conclude that the company program initiated in 1964 brought about an increase in the job satisfaction of these female clerical workers, and that this increase led to a reduction in turnover in 1966. However, since this was not a true experimental study (Campbell & Stanley, 1966, refer to it as a "pre-experimental" design which has confounded extraneous variables), several alternative explanations must be considered.

One explanation which could account for the decrease in turnover without attributing it to increased satisfaction is history. A change in the labor-market conditions from 1964 to 1966 may have occurred. It could have been much easier to get a job in 1964 than it was in 1966. If a girl was dissatisfied with her job, she could quit and easily find another job. However, an examination of the condition of the labor market for these 2 yr. indicates little change. One would expect that changes in the labor market would be reflected in the turnover rates of the other companies located in Montreal. The average annual turnover rates for the other 15 large companies in Montreal were 18.4%, 20.2%, and 20.0% in 1961, 1962, and 1963. These turnover rates have not decreased during 1964, 1965, and 1966. The average turnover rate among these 15 companies was 24.3% in 1966. An explanation for the decreased levels of turnover based on changes in the labor market conditions can be rejected.

One could explain the increases in job-satisfaction scores from 1964 to 1966 as a regression to the mean phenomenon. However, this argument loses potency when the percentile equivalents of the 1964 and 1966 satisfaction scores are examined. Only the nonsignificant change in work satisfaction (forty-fourth to forty-seventh percentile) and the change in satisfaction with co-workers (thirty-ninth to forty-seventh percentile) would fit this explanation. Futher, the change in turnover rate should not be attributed to a regression phenomenon since the turnover rate had been stable for over 10 yr. For the 5 yr. preceding the study the turnover rate fluctuated between 30.0% and 31.3%. To argue that the rate suddenly regressed to 12% stretches the notion of statistical regression.

Another explanation which could account for the results of this study is that there were changes in the composition of the female clerical workers employed at this company from 1964 to 1966. A certain pattern of these employee characteristics might predispose the worker to remain on the job and also to be more satisfied with her job. This pattern could be more frequently found among the 1966 clerical staff. If this explanation is correct then one would expect to find differences between the composition of the 1964 and 1966 clerical staff. However, a comparison of the compositions of the two staffs in terms of age, job level, years of education, and mother tongue revealed no such differences.

It would also be difficult to postulate random factors which could explain the particular pattern of results obtained. To argue that one set of uncontrolled factors accounted for the large increases (t ratios of 10.27 and 10.38) in pay and promotion satisfaction, which were the two job aspects concentrated on in the program, and small increases in the three areas not stressed and that other uncontrolled factors simultaneously led to decreased turnover strains credibility.

Other problems of internal validity resulting from this type of study design do not seem particularly relevant. Factors such as motivation, instrument decay, mortality, or selection would not seem to account for the findings.

The final remaining weakness of this design would be in the use of reactive measures (Campbell & Stanley, 1967; Webb, Campbell, Schwartz, & Sechrest, 1966). However, if the simultaneous reduction in turnover and increase in satisfaction are due purely to obtrusive measurement in combination with an otherwise ineffective program, then the millennium has indeed arrived. For if this explanation were true, then a combination of obtrusive measurement and any poorly designed, inappropriate program will solve turnover problems. This does not seem reasonable.

The major remaining question has to do with the relationship between job satisfaction and turnover under the conditions of reduced turnover rates. With a turnover rate of 30% in 1964, a multiple correlation of .34 was obtained. With a turnover rate of 12% in

1966, a multiple correlation of .45 and a biserial correlation of −.46 were obtained. The expectation was that a lower relationship between satisfaction and turnover would be obtained with reduced turnover levels. These findings indicate that the difference in the multiple correlation is nonsignificant. Even with a turnover rate of 12% a significant amount of the variance in individual termination decisions is attributable to differences in satisfaction.

REFERENCES

CAMPBELL, D. T., & STANLEY, J. C. *Experimental and quasi-experimental designs for research.* Chicago: Rand-McNally, 1966.

COOLEY, W. W., & LOHNES, P. R. *Multivariate procedures for the behavioral sciences.* New York: Wiley, 1962.

HULIN, C. L. Job satisfaction and turnover in a female clerical population. *Journal of Applied Psychology,* 1966, **50**, 280–285.

HULIN, C. L., & SMITH, P. A. An empirical investigation of two implications of the two-factor theory of job satisfaction. *Journal of Applied Psychology,* 1967, **51**, 396–402.

QUINN, R. P., & KAHN, R. L. Organizational psychology. *Annual Review of Psychology,* 1967, **18**, 437–466.

VROOM, V. *Work and motivation.* New York: Wiley, 1964.

WEBB, E. J., CAMPBELL, D. T., SCHWARTZ, R. D., & SECHREST, L. *Unobtrusive measures.* Chicago: Rand-McNally, 1966.

(Received March 20, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 2, 127–132

ACHIEVEMENT VALUES, SOCIAL DESIRABILITY, AND ENDORSEMENT OF TRAIT NAMES ON THE BERDIE CHECK LIST

BISHWA NATH MUKHERJEE [1]

Nagpur University, Nagpur, India

Using a forced-choice test of achievement motivation (v Ach), called the Sentence Completion Test (SCT), 100 Ss were classified into High v Ach and Low v Ach groups. All Ss took the Marlowe-Crowne Social Desirability scale and the Berdie Adjective Check List. It was found that Ss of High v Ach group more frequently endorsed achievement-related adjectives of the checklist. The correlation between SCT scores and frequency of endorsing achievement-related adjectives turned out to be significant even after the effects of social desirability and acquiescence were partialed out. The correlation between the later 2 response-set variables was found to be high. Implications of the results are discussed.

The adjective checklist method has recently aroused some interest among workers engaged in personality research. The adjective checklist has some advantages over some other self-report techniques of personality assessment. It takes only a few minutes for most *Ss* to check the list even though it may contain as many as 200 items. A distinct advantage of this method is that *Ss* usually are not reluctant to perform this kind of task. According to Gough (1955), the adjective checklist method has three advantages. As with the association technique, a wide segment of personality can be readily measured. The familiar words in the list provide a meaningful task for the respondent. Lastly, the endorsement of various items assures analytic ease and may stimulate research workers to derive various scoring keys from a particular checklist such as Gough's (1952) Adjective Check List (ACL).

Heilbrun (1958) developed 15 rationally derived need scales for the ACL. In one of his validational studies (Heilbrun, 1959) he found that high-achieving *Ss*, as defined by relatively high grade-point averages (GPAs) with estimated intelligence held constant, endorsed more frequently those statements which reflected need for achievement. However, in this investigation no attempt was made to control the social desirability factor (Edwards, 1957), although Heilburn (1958) himself had demonstrated earlier that frequency of endorsement of adjective need clusters is highly related to judged social desirability of the adjective characteristics of the need. Edwards (1953) has reported a high positive correlation ($r = .87$) between social desirability value of a trait name and the probability that the trait will be endorsed. Quite recently, there have been several empirical and theoretical papers directed specifically toward the problem of possible relationships between social desirability and endorsement of trait names of closely related self-descriptive statements from *Q*-sort pools and inventories like the Minnesota Multiphasic Personality Inventory (MMPI). A review of such studies will be found in Edwards (1957). Jackson and Messick (1961) found that nearly three-quarters of the common variance and more than half of the total variance in scores on MMPI could be attributed to social desirability and acquiescence. According to Edwards, Walsh, and Diers (1963), "The probability of a True response to a personality item is a linear increasing monotonic function of the social desirability scale value of the item [p. 255]." Kenny (1956) determined the mean social desirability rating (SD) based on grouped data for each of the 25 trait-descriptive adjectives and found that the SD values correlated as high as .81 with the probability of item endorsement established on the basis

[1] The author would like to thank Chandra Kishore Pd. Sinha for collecting part of the data reported in this paper.

of self-rating obtained from a separate sample.

The data reported above strongly suggest the need to control the SD factor in studies designed to test the hypothesis of differential patterns of item endorsement by groups classified in terms of a particular personality dimension. Groesbeck (1958), for example, found that persons who were high in both n Achievement and n Affiliation scales, rated themselves significantly higher on the Effective Intellectual Functioning scale (developed by Gough from MMPI items) than groups having a different configuration of the above two motives. Martire (1956) found that among *Ss* classified according to strength and generality of achievement motivation, people with low need for achievement (n Ach) endorsed to a lesser extent a number of trait names related to achievement, sociability, emotional maturity, etc., than *Ss* who obtained high n Ach scores whether achievement-related cues were present or not. Martire's (1956) data further suggest that "each group tends to see as characteristic, to some degree, those traits which they consider important [p. 370]." To what extent the reported differences between the criterion groups with respect to self-description are free from the effects of social or personal desirability of the trait names endorsed is a problem that needs systematic study. A possible difference due to acquiescent response set or "Yes-saying" has also not been assessed so far. Any attempt to determine the relationship between endorsement frequency of self-descriptive adjectives and scores on a particular personality dimension would have to be controlled both for SD and acquiescence.

The present investigation is a preliminary effort in this direction. The primary purpose of the study was to test the hypothesis that people with high verbalized need for achievement (v Ach) tend to endorse more achievement-related trait names than persons with low v Ach even after the need to present oneself in a socially approved fashion and acquiescent response set are held constant.

METHOD

Subjects

The *Ss* used were 60 undergraduate and 40 postgraduate students between the ages of 18 and 25 yr. The sample consisted of 37 males and 63 females studying in different colleges and graduate departments of Patna University, Patna, India during the 1964–65 session. The mean age of the sample was 20.39 yr. The *Ss* were all unpaid volunteers.

Instruments

The adjective checklist used in this investigation was the Femininity Adjective Check List (BACL) developed by Berdie (1959). This checklist contains a total of 148 adjectives printed in two columns with space next to each item for checking if *S* thought that particular adjective applied to him. Berdie (1959) has derived a femininity scale on the basis of *Ss*' endorsement of 61 items included in the list of 148 adjectives. The test-retest reliability of the femininity scale, as reported by Berdie (1959), is .81 for a sample of 95 men. Out of these 148 items of BACL, only 21 adjectives were regarded unanimously by 10 graduate students in psychology to be directly related to different aspects of achievement motivation. Endorsement of each of these 21 items by *S* was credited with 1 point.

The v Ach scores were obtained from the Sentence Completion Test (SCT) which was developed by Mukherjee (1964) to assess specifically consciously recognized achievement values among American college students. The SCT has also been found usable for Indian college students (Mukherjee & Verma, 1966) and has recently been used in a number of studies conducted in India. Details concerning its construction, administration, and various psychometric properties are reported by Mukherjee (1964). Basically, the SCT consists of 50 forced-choice triads (one item reflecting achievement-related sentiment and the other two pertaining to other aspects of manifest needs), selected in such a manner as to minimize the SD factor. Examples of the forced-choice triads included in SCT appear elsewhere (Mukherjee, 1968). The *S*'s score on SCT is the number of times he has chosen a statement reflecting any 1 of the following 10 aspects of achievement motivation: (*a*) hope of success, (*b*) fear of failure, (*c*) high standard of excellence, (*d*) sense of competition, (*e*) optimism, (*f*) perseverance, (*g*) concern for creative work, (*h*) preference for difficult and challenging tasks, (*i*) interest in making future plans, and (*j*) identification with a successful authority. The test-retest reliability of the SCT after an interval of nearly 2 mo. has been found in various studies to range from .71 to .83. The test-retest correlation has been found to be almost .75 for a sample of 100 Indian male graduates to whom the test was readministered after a month and a half (Mukherjee & Sinha, 1967). Previous studies (Mukherjee, 1964, 1965a, 1965b, 1965c, 1968) have demonstrated the concurrent and construct validities of the SCT as a measure of v Ach. A person high in v Ach, as measured by the SCT, is one who by word habitually indicates a desire to compete successfully with a standard of excellence. There is

some indirect evidence that the SCT also measures to some extent the degree of fear for failure (Mukherjee & Sinha, 1967; Pande & Botkawar, 1967).

The tendency to respond in a socially approved manner was measured by a very slightly modified form of the Marlowe-Crowne Social Desirability scale (M-C *SD* scale). The M-C *SD* scale (Crowne & Marlowe, 1960) consists of 33 items which were so designed following the Lie scale approach that they do not have any psychopathological implications. The scale is supposed to measure a "need for social approval." Scores on this scale have been correlated with methodologically independent, objective, life-situational or laboratory measures of personality in a number of studies reviewed recently by Crowne and Marlowe (1964), and the data suggest that the approval motive measured by the M-C *SD* scale emanates from a more basic need to maintain and defend an idealized version of the self (Crowne & Marlowe, 1964, p. 190).

The SD scale administered to *Ss* of the present investigation was slightly different from the original M-C *SD* scale in the wordings of only two items. The only major modification was with respect to Item 27 which now reads "I never act unless I have fully considered the possible consequences." The item in the original form appears as "I never make a long trip without checking the safety of my car." There was also a slight change in the wording of Item 29. Instead of "the urge to tell someone off," the modified version used "the urge to rebuke someone." The items of the modified form of the M-C *SD* scale have been found to be internally consistent (Mukherjee, 1967).

Procedure

Mimeographed copies of the SCT were distributed in class to all *Ss*. The *Ss* were urged to fill out the SCT answer sheets in their leisure time and return them within a week. All *Ss* turned in the answer sheets before time. They filled out the questionnaire in leisure time in accordance with the instructions contained on the first page of the booklet. The SCT answer sheets were scored manually following a predetermined scoring key developed on the basis of the discrimination index calculated for each item. If *S* encircled the alternative of a particular triad which had the highest discrimination index, he received 1 point for that item of the SCT, otherwise zero. Fifty *Ss* who scored 20 or above in the SCT were classified as High v Ach group and the remaining scoring 19 or less formed the Low v Ach group. There were 30 females in the High group and 33 in the Low group. The same 100 *Ss* who made up the High and Low groups also completed the Berdie Adjective Check List (BACL) approximately 3 wk. after they filled out the SCT. They were instructed to check those adjectives which they considered most descriptive of themselves. They were urged to be candid and honest in their self-evaluation. After 3 wk., all *Ss* were given the modified M-C *SD* scale with standard instructions, that is, the directions suggested by Crowne and Marlowe (1964, p. 23).

In the present study, "acquiescence" was conceptualized as a general tendency to agree but was measured in terms of the frequency of accepting many heterogeneous characteristics as descriptive of the self. The method used to assess acquiescent response set or "yes-saying" was to sum the number of adjectives endorsed by each *S*. For each *S* a frequency count was also made of the total number of achievement-related adjectives endorsed, and this measure was correlated with scores on the M-C *SD* scale and the SCT.

RESULTS

The relative frequency (proportion) of endorsement for each of the 21 achievement-related trait names is given separately in Table 1 for the two criterion groups classified on the basis of SCT score. Phi correlations estimated for each of these items following a nomograph given in Guilford (1954, p. 431)

TABLE 1

PROPORTION OF SUBJECTS IN HIGH AND LOW V ACH GROUPS ENDORSING VARIOUS ACHIEVEMENT-RELATED TRAIT NAMES AND OTHER RELATED STATISTICS

Adjective	Proportion of endorsement		Phi coefficient	Z statistic
	High v Ach	Low v Ach		
Ambitious	.74	.52	.26	2.27*
Assertive	.50	.28	.20	2.25*
Bold	.58	.40	.17	1.82
Capable	.60	.26	.34	3.43**
Confident	.64	.32	.33	3.30**
Determined	.50	.28	.19	2.22*
Enterprising	.42	.18	.25	2.62*
Forceful	.16	.04	.20	2.00*
Foresighted	.34	.16	.22	2.08*
Imaginative	.88	.62	.18	1.98*
Independent	.38	.16	.25	2.47*
Industrious	.44	.22	.28	2.35*
Initiative	.40	.18	.21	2.42*
Painstaking	.48	.22	.27	2.72**
Persevering	.56	.28	.27	2.83**
Serious	.56	.24	.31	3.26**
Sincere	.84	.52	.35	3.64**
Spontaneous	.54	.28	.25	2.65**
Steady	.76	.48	.27	2.88**
Straightforward	.64	.34	.26	3.04**
Vigorous	.18	.10	.08	1.15

* $p < .05$.
** $p < .01$.

TABLE 2

PRODUCT-MOMENT INTERCORRELATIONS AMONG MEASURES OF v ACH, SOCIAL DESIRABILITY, ACQUIESCENCE, AND FREQUENCY OF ENDORSING ACHIEVEMENT-RELATED TRAIT NAMES

Variable	Endorsement of achievement-related adjectives	Total no. endorsements	M-C *SD* scale
Sentence Completion Test	.56	.09	.15
Frequency of endorsement of achievement adjectives		.17	.20
Total no. endorsements of adjectives			.67

Note.—$N = 100$. $r = .254$ required for significance at the .01 level. For .05 level, the required value is .195.

appear in the third column of Table 1. A critical ratio test based on Z statistic was also run for each item in order to test the significance of the differences between two independent proportions of item endorsement. The results of the test of significance (Z statistics) are shown in the last column of the same table.

In Table 2 are given the product-moment intercorrelations among the major variables of the present study. From the intercorrelations, a few first-order partial correlations were also computed and they are as follows: $r_{12.3} = .55$, $r_{12.4} = .55$, $r_{13.4} = -.014$, $r_{23.4} = .049$, the first two of which are significant at the .01 level. In order to determine the intrinsic relationship between achievement values (v Ach) and frequency of endorsing achievement-related traits, a second-order partial correlation was computed in which the effects of acquiescence as well as social desirability were eliminated. The obtained coefficient $r_{12.34}$ turned out to be .56 which is significant at the .01 level.

DISCUSSION

The main purpose of the study was to examine whether or not persons classified as having higher achievement values (v Ach) endorse achievement-related traits more frequently than those who were found to have a low achievement need. The findings of the investigation clearly demonstrate that for purposes of self-description the High v Ach group chose more frequently the following: (*a*) ambitious, (*b*) assertive, (*c*) capable, (*d*) confident, (*e*) determined, (*f*) enterprising, (*g*) forceful, (*h*) foresighted, (*i*) gracious, (*j*) imaginative, (*k*) independent, (*l*) industrious, (*m*) initiative, (*n*) painstaking, (*o*) persevering, (*p*) rational, (*q*) serious, (*r*) sincere, (*s*) spontaneous, (*t*) steady, (*u*) straightforward, and (*v*) thoughtful. The *S*s of the Low v Ach group described themselves more frequently as "slow" and "submissive" than *S*s of the High v Ach group.

In a previous self-rating study, Mukherjee (1965c) also found that *S*s having high achievement values tended to rate themselves significantly higher on items pertaining to perseverance, industriousness, and diligence. No significant differences were found with respect to dimensions such as sociability, intelligence, emotional stability, etc. The findings of the present study are therefore much in line with the previous results. It is generally believed that persons with high need for achievement tend to be independent and this is supported by the study reported by Crandall, Katkovsky, and Preston (1960). Winterbottom (1958) has also reported some data which suggest that high achievement motivation is associated with early maternal encouragement for independence training. Such a relationship may also be true for achievement values (v Ach). Thus, the higher frequency of endorsement on the part of the High v Ach group of trait names such as independence, initiative, and spontaneity as found in the present study indicates more or less the same results established in the area of need for achievement (n Ach).

It is interesting, however, to note that the differences between the two criterion groups did not turn out to be statistically significant so far as their endorsement of the following trait names is concerned: active, anxious, bold, conventional, dependent, vigorous, etc. Mukherjee and Sinha (1967) reported a significant negative relationship between v Ach and manifest anxiety. It is possible that the connotations attached to these

adjectives by the *Ss* of the present study may be at variance with the conceptual meaning of the respective dimensions employed in deriving the various tools used in the studies reported above.

The first important observation emerging out of the present study is that the persons having high scores on the SCT endorsed significantly more of the total 21 achievement-related items of the checklist than did those who scored low on the SCT. The correlation between SCT scores and frequency of endorsing achievement-related items also turned out to be significant even after the effects of both acquiescence and SD were partialed out, as was hypothesized.

In the light of the present confusion over what it is that personality assessment instruments are measuring, it is relevant to question whether the significantly higher endorsement of achievement-related traits by the High v Ach group reflected basic personality differences or were merely artifacts of response styles. Christie and Lindauer's (1963) recent review points out that most existing scales and inventories are affected by both SD and tendency to agree. The results of the present study clearly show that acquiescent response set does not seem to strongly influence scores either on the BACL or on the SCT. In so far as SD is concerned, it has already been reported that the SCT and the M-C *SD* scale are essentially uncorrelated (Mukherjee & Sinha, 1967). The correlation between the two variables also turned out to be insignificant in the present study. It seems then that the attempt to minimize the SD factor in the process of constructing the SCT has not been futile.

Thus, one implication of the findings of the present study is that achievement-oriented self-perception and acquiescence set are not highly related. The study also reveals that the correlation between SD and tendency for "yes-saying" is high enough (cf. Solomon & Klein, 1963) to support the contention of Webster (1962) that "these two 'dimensions' are functionally integrated into a larger syndrome within personality [p. 789]."

The results also indicate that, as a group, the achievement-oriented assessments college students make of themselves are not very different from the ones made by an objective test of personality. This suggests that the BACL is an instrument sensitive enough to differentiate between those with high achievement values and those with low achievement values.

The present study also affords additional evidence of the construct validity of the SCT as a measure of achievement orientation by relating individual differences in value achievement to preference for endorsing achievement-related adjectives on the Berdie Check List. The findings in general support the hypothesis proposed: that what has been described as typical achievement value is a significant variable affecting one's preference with regard to self-descriptive trait names having achievement connotations. In conclusion, the results indicate that the forced-choice measure of verbalized need for achievement called SCT can not only successfully isolate the presumed differential strength of achievement values in groups but can also successfully predict the probability of endorsement of achievement-related trait names for a group even when the SD factor operates.

REFERENCES

BERDIE, R. F. A femininity adjective check list. *Journal of Applied Psychology*, 1959, **43**, 327–333.

CHRISTIE, R., & LINDAUER, F. Personality structure. *Annual Review of Psychology*, 1963, **14**, 201–230.

CRANDALL, V. J., KATKOVZKY, W., & PRESTON, A. Parent's attitudes and reactions and the development of children's achievement behaviors. Paper presented at a meeting of the American Psychological Association, Chicago, September 1960.

CROWNE, D. P., & MARLOWE, D. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 1960, **24**, 349–354.

CROWNE, D. P., & MARLOWE, D. *The approval motive*. New York: Wiley, 1964.

EDWARDS, A. L. The relationship between judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, 1953, **37**, 90–93.

EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.

EDWARDS, A. L., WALSH, J. A., & DIERS, C. J. The relationship between social desirability and internal consistency of personality scales. *Journal of Applied Psychology*, 1963, **47**, 255-259.

GOUGH, H. G. *The Adjective Check List*. Berkeley: University of California Press, 1952.

GOUGH, H. G. *Reference handbook for the Gough Adjective Check List*. Berkeley: University of California Institute of Personality Assessment and Research, 1955. (Mimeo)

GROESBECK, B. L. Toward description of personality in terms of configuration of motives. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.

GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.

HEILBRUN, A. B., JR. Relationship between the Adjective Check-list, Personal Preference Schedule, and desirability factors under varying defensiveness conditions. *Journal of Clinical Psychology*, 1958, **14**, 283–287.

HEILBRUN, A. B., JR. Validation of a need scaling technique for the adjective check list. *Journal of Consulting Psychology*, 1959, **23**, 347–351.

JACKSON, D. N., & MESSICK, S. J. Acquiescence and desirability as response determinants on the MMPI. *Educational and Psychological Measurement*, 1961, **21**, 771–790.

KENNY, D. T. The influence of social desirability on discrepancy measures between real self and ideal self. *Journal of Consulting Psychology*, 1956, **20**, 315–318.

MARTIRE, J. G. Relationships between the self concept and differences in the strength and generality of achievement motivation. *Journal of Personality*, 1956, **24**, 364–375.

MUKHERJEE, B. N. Development of a forced-choice test of achievement motivation. Final Report. Project No. S-113, United States Office of Education and Indiana University, Department of Psychology, Bloomington, August 1964.

MUKHERJEE, B. N. Achievement motivation and goal-setting behaviour in the classroom. *British Journal of Educational Psychology*, 1965, **35**, 286–293. (a)

MUKHERJEE, B. N. A forced-choice test of achievement motivation. *Journal of Indian Academy of Applied Psychology*, 1965, **2**, 85–92. (b)

MUKHERJEE, B. N. Self perception and achievement motivation. *Journal of Psychological Researches*, 1965, **9**, 109–114. (c)

MUKHERJEE, B. N. A cross-validation of the Marlowe-Crowne Social Desirability scale on an Indian sample. *Journal of Social Psychology*, 1967, **72**, 299–300.

MUKHERJEE, B. N. Achievement values and scientific productivity. *Journal of Applied Psychology*, 1968, **52**, 145–147.

MUKHERJEE, B. N., & SINHA, B. N. Some relationships between anxiety level and need for achievement. *Journal of Indian Academy of Applied Psychology*, 1967, **4**, 11–20.

MUKHERJEE, B. N., & VERMA, S. K. A cross-cultural comparison of judgments of social desirability for items of a forced-choice scale of achievement motivation. *Journal of Social Psychology*, 1966, **69**, 337–338.

PANDE, C. G., & BOTKAWAR, J. M. Relationship between achievement motivation, extraversion and neuroticism. *Journal of Indian Academy of Applied Psychology*, 1967, **4**, 54–58.

SOLOMON, L., & KLEIN, E. The relationship between agreeing response set and social desirability. *Journal of Abnormal and Social Psychology*, 1963, **66**, 176–178.

WEBSTER, H. Acquiescence, social desirability and inhibition reflected by "response set" scales. *Psychological Reports*, 1962, **10**, 789–790.

WINTERBOTTOM, M. R. The relation of need for achievement to learning experience in independence and mastery. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.

(Received March 20, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 2, 133–138

EFFECTS OF SIGNAL AND STIMULUS RATES ON VIGILANCE PERFORMANCE

HARVEY A. TAUB [1] AND FRANCIS H. OSBORNE

Syracuse Veterans Administration Hospital

2 experiments investigated the relative effects of critical signal rate, normal noncritical signal rate, and critical signal probability on detection in a visual vigilance task. In all conditions Ss had to monitor a clock display for a period of 54 min. The results of both experiments suggested that normal signal rate may be of greater importance in a vigilance situation than either critical signal rate or critical signal probability. In general, both percentage of detections and percentage of stimuli falsely responded to were inversely related to normal signal presentation.

Although numerous experiments have indicated that probability of detection is directly related to critical signal rate, recent studies of vigilance behavior have suggested that additional rate variables, such as noncritical signal and total stimulus rates, may be of equal or even greater importance in determining vigilance performance. Colquhoun (1961) observed that many of the studies which found a direct relationship between performance and frequency of critical signal presentation (Jenkins, 1958; Kappauf & Powe, 1959) had utilized tasks in which an increase in the number of critical signals was simultaneously accompanied by a decrease in normal or noncritical signals. Accordingly, Colquhoun investigated the separate effects of both critical and normal signal rates and concluded that detection efficiency was determined by the probability that a stimulus will be a critical signal, that is, number of critical signals divided by the total number of stimuli presented. Subsequent studies (Jerison & Pickett, 1964; Jerison, Pickett, & Stenson, 1965; Johnston, Howell, & Goldstein, 1966) have confirmed the general findings of Colquhoun by obtaining differences in performance which were directly related to critical signal probability. On the other hand, the results from a recent study by Jerison (1965) have suggested that the main variable affecting performance may be the number of normal signals (noncritical) and not critical signal probability. Jerison's (1965) findings indicated that under certain conditions percentage of detections decreased as the number of noncritical signals increased, even where critical signal probability was equal.

Since each of the above studies (Colquhoun, 1961; Jerison et al., 1965) utilized only two or three groups and thus only two comparison levels of each of the relevant variables, the discrepancy between findings might be due to the particular levels which were investigated. Thus, the present experiments were designed to provide further insight into the relative effects of critical signal rate, normal signal rate, and critical signal probability. This was accomplished in Experiment I by investigating the effects of critical signal and total stimulus rates at various signal probabilities and in Experiment II by using a constant critical signal probability level derived from several combinations of critical and noncritical signal rates.

EXPERIMENT I

Method

Subjects. Data are reported on a total of 72 undergraduate males who participated as part of the requirements of the introductory course in psychology at Syracuse University. Data from nine additional *Ss* were discarded due to equipment failure ($N = 4$), not responding ($N = 3$), or continuous random responding ($N = 2$).

Apparatus. The visual display was a Type S-1 Standard Electric clock with a 1-rps sweep rate which had been modified as follows: (*a*) the instrument face was reversed and painted white; (*b*) clock face diameter was reduced from 3½ to 2¾ in. by painting the outer rim of the glass display surface white; and (*c*) the black sweep hand was painted white toward the center of the display, leaving only a ½-in. black portion at the tip. All three modifications were based on preliminary

[1] The author is also at the State University of New York, Upstate Medical Center, Syracuse.

testing and were designed to increase the difficulty of the vigilance task.

As in other clock-type vigilance tasks, pulses were presented at fixed rates which allowed the sweep hand to move or jump a specified distance. The normal or noncritical signal consisted of a 9° movement, while the critical signal to be detected by *S* had a sweep-hand movement of 20°. Thus, *S*'s task was to detect a "double jump" which was actually 2.2 times the distance of the normal noncritical signals. Detection responses were to be made on a telegraph key located directly in front of *S*.

Two identical clocks were modified and placed in adjoining rooms to allow for simultaneous testing of two *Ss*. Each clock was positioned at one end of a table with an acoustically tiled wall behind it. The *S* was located at the other end of the table approximately 4½ ft. from the clock. Further, to reduce possible noise and vibration transmission, each clock was placed upon acoustic-absorbing material. Finally, to insure complete acoustic isolation and a uniform acoustic environment, *Ss* wore earphones and received 75 db. of white noise throughout the testing session.

All programming and recording equipment were located in an adjoining control room. The rate of stimulus presentation was controlled by Grason-Stadler Programming modules, while critical signal rate and onset were controlled by a Gerbrands tape programmer. An Esterline Angus 20 pen Model recorder and a Grason-Stadler Print-Out counter were used to record each signal and each *S*'s response. Formal instructions and white noise were prerecorded and played back through *Ss*' earphones by a Wollensak tape recorder.

Signal schedules. Stimuli were presented at two fixed rates of 30 and 60 stimuli per minute (i.e., normal plus critical signals), providing a total of 1,620 and 3,240 stimuli, respectively, for the 54-min. session. The preprogrammed film tapes provided that 27, 54, or 108 of the stimuli presented in each test session were critical signals while the rest were normal signals. Each program tape was set up with a time base of 2 min., such that within each critical signal rate the same number of signals were presented in each 2-min. block, that is, one, two, or four critical signals per 2 min. Critical signals were programmed randomly within the restrictions imposed by the 2-min. blocks. Table 1 presents the six combinations devised from the three critical signal and the two total stimulus rates and, for each condition, further illustrates the probability of a stimulus being a critical signal.

Procedure. The *Ss* were randomly assigned to 6 groups with 12 *Ss* per group. Each group served under a different combination of total stimulus rate and critical signal rate. All *Ss* served for 54 min., thus providing a measure of performance for three 18-min. time blocks.

At the beginning of the testing session, *Ss* were given informal instructions which drew attention to the clock and key and which provided stimulus samples. The *Ss* continued to receive sample stimuli until they could discriminate between the normal and critical signals. At this time *Ss* were told to put on their earphones and were given the formal pretaped instructions for the vigilance task. The formal taped instructions provided further examples of the normal and critical signals and instructed *Ss* that their job was to press the telegraph key as soon as a critical signal was detected. The *Ss* were further informed of the white noise and were requested to remain seated and to continue the task until *E* entered the room to terminate the session.

All watches were removed before the testing session began so *Ss* did not receive any time cues, although they did know that the session would last "about an hour."

TABLE 1

CRITICAL SIGNAL PROBABILITY FOR EACH OF THE 6 EXPERIMENTAL CONDITIONS

Total stimuli	Critical signals		
	27	54	108
1620	.0166	.0333	.0666
3240	.0083	.0166	.0333

Results

A single response-time criterion was used for all groups even though the duration of the interval between successive stimuli was inversely dependent upon rate of stimulus presentation. Since *Ss* tended to pace their responses to the off-intervals, the single criterion was based upon the duration for the groups with the slowest stimulus rate. Thus, a response to a critical signal was considered as being correct only if it occurred within a 2-sec. interval following signal presentation. All responses which occurred outside of this interval were considered as false detections.

Correct detections. The number of critical signals detected by each *S* during each of the three 18-min. time periods was converted to percentage correct by dividing the number of signals detected by the number of signals presented. A Hartley's test of homogeneity of variance (Winer, 1962) was run using the between-groups effects of critical signal rate and total stimulus rate and indicated that the assumption of homogeneity was tenable and that all groups could be run in a single analysis of variance. The analysis of variance of the data indicated that the only significant effects were total stimulus rate, F (1/66)=

14.25, $p < .01$, and time periods, F (2/132) $= 36.62$, $p < .01$. No consistent differences in performance were attributable to critical signal rate or to the interactions between the variables.

Figure 1 presents mean percentage of correct detections for each of the six groups during each of the three 18-min. time periods. The main effect of time periods is clearly shown by the fact that all groups made less detections in each successive period. Similarly, the data underlying the significant effect of total stimulus rate indicate that performance was poorer with the faster rate of stimulus presentation as compared to the slower rate. The lack of interaction between these variables suggests that, in general, the increase in total number of stimuli did not affect the rate of vigilance decrement but, in effect, lowered performance in all time periods.

Of further relevance to the evaluation of the various rate parameters in vigilance performance are the findings that, although in the predicted direction, the effects due to critical signal rate were not sufficiently large or consistent to achieve statistical significance. These findings plus the significant effect of total stimulus rate suggest that the rate of noncritical or normal signal presentation may be the most important of the variables under question. That is, the combined results indicate that a fourfold increase in critical signals and signal probability with a relatively small decrease in normal signals produced no consistent effect, whereas holding critical signals constant and varying the number of normal signals produced a significant effect. Finally, even in conditions of equal probability (see Table 1), differences appeared as a function of normal signal rate.

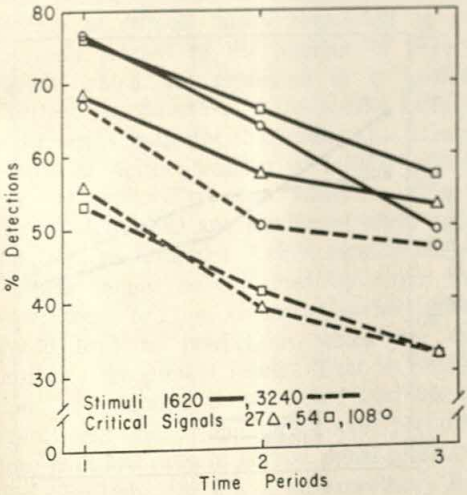


FIG. 1. Percentage of detections as a function of critical signals, total stimuli, and 18-min. time periods.

False detections. The number of false detection or commission responses varied considerably with individual *S*s making as few as 1 or as many as 320 commissive errors. Further, 25% of the *S*s made 65% of the total false detections. Thus, high variability obscured any general between-group trends. The only trend which was consistent was that commissive errors decreased with time, F (2/132) $= 16.72$, $p < .01$.

Conversion of the false-report data into percentages (i.e., number of commissive errors divided by number of normal stimuli that could be responded to with the 2-sec. detection criterion) further suggested that the proportion of normal signals responded to was inversely related to total stimulus rate, F (1/66) $= 9.46$, $p < .01$. Mean percentages of responses were 5.68% and 2.57%, respectively, with the slow and fast stimulus rates. No other trends were consistent except the decrement with time (as noted above). In general, the results closely parallel the findings for percentage of correct detections; that is, the general performance trends are identical with a greater percentage of normal signals being responded to in those conditions which also led to the highest levels of correct detections.

EXPERIMENT II

The findings of Experiment I suggested that noncritical signal rate may be of greater importance in a vigilance situation than either critical signal rate or critical signal probability. Since all three factors could not be compared directly in Experiment I, a second experiment was run to further evaluate the relative effects of each of the variables. This was accomplished at various stimulus rates by simultaneously varying the number of normal and critical signals in a manner which maintained a single critical signal probability

level. Three predictions could be made with this design: (1) if probability level is the most important factor, then rate of stimulus presentation should have no effect upon performance; (2) if noncritical signal rate is the most important variable, then detection performance should vary inversely with rate of stimulus presentation; and (3) if critical signal rate is of greatest importance, then performance should vary directly with stimulus presentation rate.

Method

The *Ss* consisted of 20 male and 20 female undergraduate students at Syracuse University. Data from seven additional students were discarded due to equipment failure ($N = 6$) or refusal to complete the session ($N = 1$).

Five male and five female *Ss* were randomly assigned to four stimulus rate groups which were derived from different combinations of normal signal rate and critical signal rate. These combinations, which will be written in the form "critical signal rate/normal signal rate," are 216/6264, 108/3132, 54/1566, and 27/783. All combinations have the identical signal probability of .033; that is, the ratio of total critical signals to total stimuli is .033 in all four combinations.

The basic apparatus, task instructions, and methods of making the stimulus tapes were identical to those used in Experiment I. In all cases, *S*'s task was to depress the telegraph key upon detection of a critical signal. Similarly, all *Ss* served for a total of 54 min.

Results

As in Experiment I, the criterion for a correct detection response was determined by the group which had the longest duration between the presentation of succeeding stimuli, that is, Group 27/783 with 810 total stimuli per 54 min. and 1 every 4 sec. Thus, in the present experiment the cutoff time was 4 sec., and all responses were counted as being correct if made within 4 sec. after a critical signal was presented. All other responses were considered as being false detections.

Correct detections. The Hartley test indicated homogeneous variances for the four experimental groups. Analysis of variance of percentage of detections indicated that stimulus rate groups, $F\ (3/32) = 6.41$, $p < .01$, time periods, $F\ (2/64) = 18.52$, $p < .01$, and the Time Periods × Sex interaction, $F\ (2/64) = 3.29$, $p < .05$ were the only significant effects. As in Experiment I, the interaction between time periods and stimulus rate was not significant.

TABLE 2

PERCENTAGE OF CRITICAL SIGNALS CORRECTLY DETECTED AND PERCENTAGE OF NORMAL SIGNALS WHICH RECEIVED COMMISSION RESPONSES

Group	Correct detections	False responses
216/6264	43.52	1.21
108/3132	44.64	1.45
54/1566	72.62	8.99
27/783	61.86	7.31

Table 2 illustrates the data underlying the findings for the stimulus rate groups and indicates that performance was inversely related to the rate at which the normal or noncritical signals were presented. More specifically, the table suggests the two slowest critical signal/normal signal combinations led to better performance than obtained with the two faster rates. Further, the lack of interaction with time periods suggests that these differences were apparent in all three 18-min. blocks. Figure 2 presents the data for the Time Periods × Sex interaction and suggests that although there was a general decrement in performance over time, the rate of decrement was greater for males than females.

False detections. The *Ss* exhibited a large

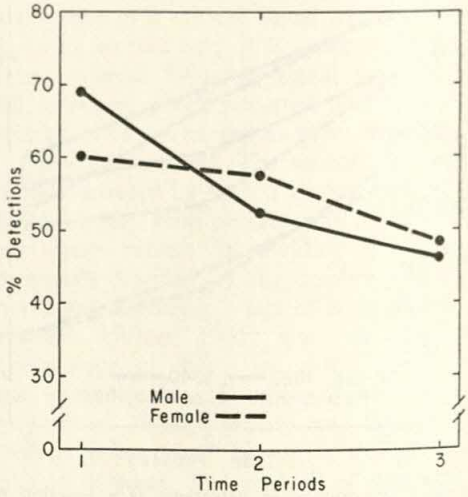


FIG. 2. Percentage of detections for male and female subjects as a function of 18-min. time periods.

amount of variability in the number of false detections, with 20% (eight *Ss*) making 65% of the commissive errors. Further, although the number of commissive errors for the 54/1,566 group was more than twice that of the other groups, this difference is attributable to the error total of one *S*, which, in fact, was equal to or greater than the total number of errors for each of the other three groups. Thus, it did not appear that the groups differed in terms of number of false detections. Finally, although number of false reports appeared to decrease with time, an analysis indicated that this decline was not significant.

Conversion of number of false reports into percentages by dividing by the total number of normal signals which could be responded to with the 4-sec. response criterion indicated that stimulus rate groups was a significant effect, $F\ (3/36) = 3.75$, $p < .05$ with this measure. That is, since the number of commissions appeared to be similar, the fewer the number of signals presented, the higher the percentage which was responded to by *S*. Table 2 indicates that the general trend of these data closely resembles that for correct detections.

DISCUSSION OF EXPERIMENTS I AND II

In general, the present results suggest that both percentage correct detections and percentage of stimuli falsely responded to are inversely related to the number of normal signals which are presented in a session. Further, the direction of the results in both experiments indicates that under the present conditions neither number of critical stimuli nor the probability of a stimulus being a critical signal had any significant effect upon detection performance. Taken together, these findings cannot be explained by either the expectancy or the arousal theories which would lead to predictions which are opposite to the present results. That is, predictions from arousal theory would lead one to expect less arousal and thus poorer performance with low rates of normal signal presentation. Similarly, the use of an expectancy hypothesis would predict, in Experiment II, overall detection performance to be directly related to critical signal rate and amount of decrement over time to be inversely related to critical signal rate. The only explanation which appears to be compatible with the present results is a decision-theory model of vigilance which suggests that detection performance is directly related to observing responses which are, in turn, dependent upon the various signals rates (Jerison, 1965; Jerison & Pickett, 1963, 1964). Although the present results are not in agreement with the suggestion that signal probability is the most effective variable, they are in agreement with the assumption that the cost of an observing response increases directly with normal signal rate, while the expected value of that response varies inversely (Jerison, 1965). If this latter assumption can be accepted then the present results for both percentage of correct and percentage of false detections can be explained by a decrease in observing responses, attention to the display, and detection responses due to an increase in the number of normal signals with a consequent increase in cost of observing and decrease in value of observing.

Although the present findings are contrary to expectations from the study by Colquhoun (1961), a number of possible differences were present in the experiments and might account for the discrepancies in results. Colquhoun's use of a slow normal signal rate was one possible difference, although Jerison (1965) found that almost all signals were detected with few normal signals regardless of the probability. A second possible reason might be due to Colquhoun's training conditions which, in contrast to the present study, provided pretest practice trials using the same signal probabilities as were experienced in the testing session. A third difference and one of possible greater importance is the fact that Colquhoun actually utilized three types of stimuli with the third being a blank display. This third type of stimulus, in effect, informed *Ss* that neither a critical nor noncritical signal had appeared and thus may have affected the nature of responses. The final and probably most likely contributor to the discrepancy is that Colquhoun's single comparison of equal-signal probabilities, which produced no significant differences in percentage of detections, was

made using 0.5 as the probability level whereas the highest probability of a stimulus being a critical signal was .06 in the present experiments and .04 in the studies by Jerison (1965). In view of the large difference in signal probabilities, it appears possible that dual processes may be present such that the normal signal rate may be the determiner of performance at low probabilities, while at high probabilities either (*a*) the probability itself determines performance, or (*b*) detection performance has reached an asymptote and is not affected by further changes in rate or probability. These factors require further research before any general conclusions may be drawn.

Finally, the effect of males versus females, although of lesser importance than the other results, suggests that females are better able to maintain performance over a long period of time. This interaction is in general agreement with the results of a study by Whittenburg, Ross, and Andrews (1956), although the exact shape of the interactions is different due to different time measurements. That is, the present study suggested that males were superior in the first 18-min. period with a reversal occurring in the second and third period, whereas Whittenburg et al. using half-hour time periods found no differences during the first hour with the females being superior to male *S*s during the last hour.

REFERENCES

COLQUHOUN, W. P. The effect of 'unwanted' signals on performance in a vigilance task. *Ergonomics,* 1961, **4**, 41–51.

JENKINS, H. M. The effects of signal rate on performance in visual monitoring. *American Journal of Psychology,* 1958, **71**, 647–661.

JERISON, H. J. Human and animal vigilance. *Perceptual and Motor Skills,* 1965, **21**, 580–582.

JERISON, H. J., & PICKETT, R. M. Vigilance: A review and re-evaluation. *Human Factors,* 1963, **5**, 211–238.

JERISON, H. J., & PICKETT, R. M. Vigilance: The importance of the elicited observing rate. *Science,* 1964, **143**, 970–971.

JERISON, H. J., PICKETT, R. M., & STENSON, H. H. The elicited observing rate and decision processes in vigilance. *Human Factors,* 1965, **7**, 107–128.

JOHNSTON, W. A., HOWELL, W. C., & GOLDSTEIN, I. L. Human vigilance as a function of signal frequency and stimulus density. *Journal of Experimental Psychology,* 1966, **72**, 736–743.

KAPPAUF, W. E., & POWE, W. E. Performance decrement at an audio-visual checking task. *Journal of Experimental Psychology,* 1959, **57**, 49–56.

WHITTENBURG, J. A., ROSS, S., & ANDREWS, T. G. Sustained perceptual efficiency as measured by the Mackworth "clock" test. *Perceptual and Motor Skills,* 1956, **6**, 109–116.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received March 22, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 2, 139–144

PARTICULARISM AND ORGANIZATIONAL POSITION: A CROSS-CULTURAL ANALYSIS [1]

LOUIS A. ZURCHER [2]

The Menninger Foundation, Topeka, Kansas

The Stouffer-Toby Role Conflict Scale, a measure of particularism (the value for institutionalized obligations of friendship), was administered to 230 employees in 13 bank branches in Mexico and the United States. Findings support the hypothesis that for both bank officers and line employees Mexicans are significantly more particularistic than Mexican-Americans, who in turn are significantly more particularistic than Anglo-Americans. The hypothesis that, by ethnic group, officers are less particularistic than line employees was partially supported. The cultural development of particularism and the potential impact of that value upon organizational behavior are discussed. Some suggestions are offered to management and to researchers concerning the importance of understanding the interactions among specific value orientations and specific organizational behaviors, and concerning the implications of such interactions for employee satisfaction and organizational integrity.

Rapid industrialization of "underdeveloped" countries, continued expansion of international business operations, and increased use of indigenous labor by foreign companies heighten the relevance of cross-cultural studies of individual behavior in complex organizations. Though such studies are relatively few, their findings invariably indicate the pervasive influence of cultural values upon, for example, the pace and pattern of company development, the emergence of management and supervisory styles, and the varieties and intensities of motivation for, attitudes toward, and alienation from work (Fayerweather, 1959; Haire, Ghiselli, & Porter, 1963; McClelland, 1961; Whitehill, 1964; Whyte, 1963; Whyte & Williams, 1963; Zurcher, Meadow, & Zurcher, 1965).

Ideally, the investigator seeking to appreciate the dynamics of cross-cultural differences in organizational behavior would himself experience the life styles and work settings of his subjects (*S*s), and then come to grips with the complex interaction between whole personalities and whole organizations. Considerably less ambitious, the purpose of this paper is to isolate a single cultural value, and to observe how it varies between and within specific work organizations.

The term "value" is taken to mean "a selective orientation toward experience, implying deep commitment or repudiation, which influences the ordering of choices between possible alternatives in action [Kluckhohn, 1961, p. 18]." Values are learned in one's parent culture, and "culture" is "the distinctive way of life of a group of people, their complete design for living [Kluckhohn, 1951, p. 86]." The value to be studied in this paper is "particularism," or the value for institutionalized obligations of *friendship*. Particularism is one component of Parsons and Shils' (1959) general theory of action and is juxtaposed with "universalism," or the value for institutionalized obligations to *society* (pp. 76–77). According to Parsons and Shils, before acting in a specific social situation the individual will be influenced by

[1] This paper is based on the author's unpublished doctoral dissertation, University of Arizona, Tucson, 1965. The research was supported by the National Science Foundation through a Cooperative Graduate Fellowship awarded to the author, and by the National Institute of Mental Health, Grant No. 5-R11-MH-544-2, Arnold Meadow, director. Special acknowledgment is made to Arnold Meadow, Salvatore Zagona, William MacKinnon, Richard Coan, Stuart Queen, and Jerry L. L. Miller of the University of Arizona for their helpful comments. Without the cooperation of John Henderson, vice-president of the Valley National Bank, Tucson, Arizona, the Banco Nacional de Mexico, Nogales, Sonora, and the Bureau of Business and Public Research, University of Arizona, this study would not have been possible.

[2] Now in the Department of Sociology, University of Texas.

a particularistic or a universalistic orientation, and that orientation will have been culturally determined (p. 79). Using varying vocabularies, several other social scientists have, based upon their comparative observations of societies in differing stages of social complexity, conceptualized dichotomies or continua bounded on one side by a predominant value for concrete interpersonal ties and on the other by a predominant value for abstract societal expectations (Becker & Meyers, 1942; Lewis, 1951; Loomis & Beegle, 1950; Miner, 1952; Redfield, 1941). Members of the less complex, more "primitive" societies are reported to tend toward particularism, members of complex societies toward universalism.

In that context, Mexico has been described as manifesting (modally) a cultural value for particularism. Parsons and Shils (1959), Lewis (1951), Redfield (1947), and Cohen (1961) report the influence of the Mexican social structure, particularly the extended kinship system, upon the individual's assessment of events in his daily life. Government, laws, and other abstract societal expectations are observed to be distant, ephemeral, and untrustworthy to the Mexican, and not sufficient basis for the evaluation of other human beings. What seems more important is the depth and significance of interpersonal ties with blood relatives, *compadres,* and friends. Another person is accepted by the individual not in terms of office held, title displayed, or any other such abstraction, but in terms of immediate interpersonal experience with the evaluating individual. The essential questions are: What kind of person is he to me? Can he be trusted?

On the other hand, the United States has been described (modally) as a universalistically oriented society. The high level of technology, "scientism," and bureaucracy, the keen emphasis upon competition, achievement, and mobility encourage more shallow interactions with other human beings and a greater dependence upon collective societal definitions of others (Buber, 1958; Fromm, 1940; Riesman, 1950). The essential questions are: What (group, company, department, etc.) does he belong to? How high up is he?

If these observations on the two societies are accurate, differences in selected cultural values, specifically particularism, should be empirically demonstrable. That the present discussion of Mexico and the United States has been in terms of *modal* value orientations admits to the fact that, for example, groups in rural America may be more particularistic than groups in highly urbanized Mexico City. Thus, as with all cross-cultural research, special effort must be made to establish comparability of samples. In this study, comparability was approached by drawing *Ss* from similar work organizations in Mexico and the United States. The structures of the work organizations were such that employees could be classified as "officer" or "line." It was therefore hypothesized that otherwise comparable Mexican officers would be more particularistic than Mexican-American officers, who would in turn be more particularistic than Anglo-American officers. A similar directional difference in particularism was hypothesized for line employees.

Since the work organization was bureaucratic in nature, the assumption would be that movement up the organizational ladder would be associated, by a selection or a socialization process, with a shift away from the particularistic value orientation. Some indications of that relationship can be found in the work of Merton (1940) and Whyte (1948). In this study it was therefore hypothesized that Mexican officers would be less particularistic than Mexican line employees, Mexican-American officers less particularistic than Mexican-American line employees, and Anglo-American officers less particularistic than Anglo-American line employees.

METHOD

Subjects

The work organization chosen as the setting for this study was the bank, because it provided an accessible and cross-culturally comparable bureaucratic organization from which to draw *Ss*. Mexican and American banks are almost identical in formal organization and operating procedures (Beteta, 1963), the former being based upon the structure of the latter.

Three experimental groups were used, the number of *Ss* for all three groups totaling 230. The

first group was composed of native-born, Spanish-speaking (unilingual) Mexican citizens employed by two bank branches in Nogales, Sonora. The second group was composed of bilingual (Spanish-English) Mexican-Americans employed by two bank branches in Nogales, Arizona (bilingualism and United States citizenship are conditions for employment in these branches). The third group was composed of English-speaking (unilingual) Anglo-Americans employed by nine bank branches in Tucson, Arizona. All 11 United States banks are branches of the same parent organization. Both of the Mexican banks are branches of the same Mexican parent organization. All employees of the banks (up to the position of vice-president and including at least one vice-president for each cultural group) who were at work on that day were administered the questionnaire. The final sample included all of these employees, with the exception of 1 Anglo-American who was removed from the Mexican-American group, and 14 Mexican-Americans who were removed from the Anglo-American group. The employees of the Nogales, Sonora, banks were all Mexicans. All three groups approximated each other for proportion of officers (about one to every six line employees), for employee longevity ($Mdn = 2$–4 yr.), and for age ($Mdn = 20$–29 yr.). The ratio of males to females was found to be disproportionate among the groups (22:16 Mexican; 14:29 Mexican-American; 66:83 Anglo-American), but subsequent analysis demonstrated sex not significantly to be related to particularism. Officers ($N = 35$) were vice-presidents, managers, assistant managers, and all other employees who had "officer" in their formal titles (i.e., loan officer, personnel officer, etc.). Line employees ($N = 195$) were tellers, clerks, typists, secretaries, including first-line supervisors (i.e., head teller, chief clerk, etc.).

Measures

As part of a questionnaire packet that contained biographical and employment history questions and some attitude toward work scales, each *S* was given the Stouffer-Toby Role Conflict Scale. The Role Conflict Scale is a four-item Guttman instrument designed to sound the degree of particularism by presenting four situations involving conflicts between obligation to a friend and more general social obligations, and forcing the individual to choose one set of obligations or the other (Stouffer & Toby, 1951). For example, Item 1 states:

> You are riding in a car driven by a close friend, and he hits a pedestrian. You know he was going at least 35 miles an hour in a 20-mile-an-hour speed zone. There are no other witnesses. His lawyer says that if you testify under oath that the speed was only 20 miles an hour, it may save him from serious consequences.
> What right has your friend to expect you to protect him?
>
> *Check one:*
>
> ——My friend has a definite right as a friend to expect me to testify to the lower figure.
> ——He has some right as a friend to expect me to testify to the lower figure.
> ——He has no right as a friend to expect me to testify to the lower figure.
>
> What do you think you'd probably do in view of the obligations of a sworn witness and the obligation to your friend?
>
> *Check one:*
>
> ——Testify that he was going 20-miles-an-hour.
> ——Not testify that he was going 20-miles-an-hour.

The response "My friend has a definite right . . ." is considered to be particularistic. The response "He has no right . . ." is considered to be universalistic. If the respondent marks "He has some right . . ." the response is scored particularistic if in the second part of the question he says he would favor the friend in action; universalistic if he says he would not favor the friend. Scale scores may vary from 0 to 4, with a score of 0 indicating a minimum and a score of 4 indicating maximum particularism. Only particularistic responses are scored, 1 point for each item.

The Role Conflict Scale was translated into Spanish for use with the Mexican *S* group. Both the English and the Spanish scales maintained Guttman scalability throughout the present study. The coefficient of reproducibility was .924.

Procedure

The mimeographed questionnaire form was administered, either before or after customer hours, to the on-the-job employees by the investigator. Prior to the administration, the employees were gathered together and told by the investigator that their anonymous responses were considered to be confidential, would not be seen by any other bank personnel, and would be used only for research purposes. The employees were then given copies of the questionnaire, and went to their own work spaces to fill them out. When finished with the forms, the employees returned them to the investigator. Members of management simultaneously filled out the forms and did not influence the line employees in any apparent way. The same procedure was used in every bank branch, and all the data were collected within a time span of 20 working days.

RESULTS

Table 1 presents the distribution of bank-officer particularism scale scores for each enthnic group. A Kruskal-Wallace one-way analysis of variance indicated that particular-

ism was significantly related to ethnic membership ($H = 35.08$, $df = 2$, $p < .0005$). Further analysis by Fisher's Exact Probability test (*S*s who scored 0 versus those who did not) and inspection of Table 1 indicate the direction of the overall results: Mexican versus Mexican-American officers ($p = .045$); Mexican versus Anglo-American officers ($p = .0004$); Mexican-American versus Anglo-American officers ($p = .008$). Thus, as hypothesized, the particularism of the bank officers is influenced by cultural background—Mexicans significantly more particularistic than Mexican-Americans, who in turn are significantly more particularistic than Anglo-Americans.

Table 2 presents the distribution of line-employee particularism scale scores for each ethnic group. As with the officers, a Kruskal-Wallace one-way analysis of variance indicated that line-employee particularism was significantly related to ethnic membership ($H = 9.92$, $df = 2$, $p < .0005$). Further analysis by the Median Test (one-tailed) and inspection of Table 2 indicate the direction of the overall results: Mexican versus Mexican-American line employees ($\chi^2 = 14.44$, $df = 1$, $p < .0005$); Mexican versus Anglo-American line employees ($\chi^2 = 24.80$, $df = 1$, $p < .0005$); Mexican-American versus Anglo-American line employees ($\chi^2 = 3.9$, $df = 1$, $p < .025$). Again as hypothesized, the particularism of the bank line employees is influenced by cultural background—Mexicans significantly more particularistic than Mexican-Americans, who in turn are significantly more particularistic than Anglo-Americans.

With ethnic groups combined, officers were demonstrated to be significantly less particularistic than line employees ($H = 3.56$, $df = 5$, $p < .05$). However, the hypothesis that officers would be significantly less particularistic than line employees for *each* ethnic group was supported only for the Anglo-Americans ($H = 10.31$, $df = 1$, $p < .0005$). Inspection of Tables 1 and 2 suggests that the distribution of Mexican and Mexican-American scores tends toward the hypothesized difference between officers and line employees, but the small *N*s, particularly for the officers, make interpretation tenuous and obscure statistical analysis.

TABLE 1

OFFICER PARTICULARISM BY CULTURAL ORIGIN

Cultural origin	Scale score					*N*
	0	1	2	3	4	
Mexican	0	2	2	1	0	5
Mexican-American	4	0	0	1	1	6
Anglo-American	21	3	0	0	0	24
Total	25	5	2	2	1	35

TABLE 2

LINE-EMPLOYEE PARTICULARISM BY CULTURAL ORIGIN

Cultural origin	Scale score					*N*
	0	1	2	3	4	
Mexican	5	5	6	7	10	33
Mexican-American	17	12	4	4	0	37
Anglo-American	82	23	10	8	2	125
Total	104	40	20	19	12	195

DISCUSSION

The hypothesis that a single value, particularism, varies cross-culturally for officers and line employees in a specific type of work organization has been supported. This finding encourages the conclusion that value orientations are central factors to be considered in the social psychology of work, but raises further questions about the *relation* of such values to measurable organizational *behavior*. Zurcher et al. (1965) report findings with the present *S* group that indicate a significant positive relation between particularism and alienation from work in the bank, which is interpreted to be a universalistic organization. The same authors point to an active social club in the Mexican-American *S* banks and suggest that those informal ties may be the result of particularistic needs. McClelland (1961) reports Mexican businessmen to manifest a significantly higher need for affiliation (a particularistic parallel) and a significantly lower need for achievement than do Anglo-American businessmen. Fayerweather (1959) observes that

Mexican businessmen seem to be more concerned with the maintenance of personal alliances than with bureaucratic efficiency and if forced to choose between those alternatives will choose the former. Whyte and Williams (1963, p. 6) conclude that the Peruvian worker "looks upon the supervisor as a total human being, without thinking about the various aspects of his job performance," and that this "fits in with the emphasis upon the person and personalism" that has been noted in Latin American culture.

Further research, certainly interdisciplinary in nature, is needed to delimit more specifically the degree to which particularism or other values vary cross-culturally and the impact they have upon organizational behavior—that is, performance, supervision, policy formulation, absenteeism, turnover, etc. Especially useful would be data illustrating the generation of conflicts in personnel expectations from disparities in value orientations, when and if such conflicts are functional or disfunctional for the individual and for the organization, and the means, if appropriate and desirable, for the amelioration of those conflicts.

The present study was not able clearly to demonstrate for each of the three ethnic groups that officers were significantly less particularistic than line employees. Insufficient officer N may account for the lack of statistically significant differences between officer and line Mexicans and Mexican-Americans—a difference which was markedly demonstrated for the considerably larger Anglo-American group. On the other hand, it may be that the value for particularism is not importantly altered by or a requisite for managerial status in the Mexican and Mexican-American banks. The data are not conclusive for either interpretation. The significant difference established in the Anglo banks supports the original hypothesis, but introduces an interesting question. What proportion of the difference in particularism can be accounted for by the screening and selecting process of promotion, and what proportion can be accounted for by the process of officer role-enactment *following* promotion?

The data and interpretations presented in this paper suggest to management in complex organizations, particularly those organizations with international branches or subsidiaries, that differences in the cultural background of personnel may yield differences in their value orientations—value orientations which may influence organizational behavior. Also, management should not assume that a set of policies, organizational expectations, or leadership styles established and proven successful in one culture will inevitably be successful in another.

The findings of this study suggest to the researcher that there is continued need for interdisciplinary studies of the interactions among specific cultural values and specific organizational behaviors.

REFERENCES

BECKER, H., & MEYERS, R. C. Sacred and secular aspects of human association. *Sociometry*, 1942, 5, 207–229.

BETETA, M. Banking laws of Mexico. *Mexico This Month*, August 1963, 7–10.

BUBER, M. *I and thou*. New York: Scribner's, 1958.

COHEN, Y. A. Patterns of friendship. In Y. A. Cohen (Ed.), *Social structure and personality*. New York: Holt, Rinehart & Winston, 1961. Pp. 353–383.

FAYERWEATHER, J. *The executive overseas*. New York: Syracuse University Press, 1959.

FROMM, E. *Escape from freedom*. New York: Farrar & Rinehart, 1940.

HAIRE, M., GHISELLI, E. E., & PORTER, L. W. Cultural patterns in the role of the manager. *Industrial Relations*, 1963, 2, 95–117.

KLUCKHOHN, C. The study of culture. In D. Lerner & H. D. Lasswell (Eds.), *The policy sciences*. Stanford: Stanford University Press, 1951. Pp. 86–101.

KLUCKHOHN, C. The study of values. In D. N. Barrett (Ed.), *Values in America*. Notre Dame: University of Notre Dame Press, 1961. Pp. 17–45.

LEWIS, O. *Life in a Mexican village: Tepoztlan restudied*. Urbana: University of Illinois Press, 1951.

LOOMIS, C. P., & BEEGLE, J. A. *Rural social systems*. New York: Prentice-Hall, 1950.

MCCLELLAND, D. *The achieving society*. Princeton: Van Nostrand, 1961.

MERTON, R. Bureaucratic structure and personality. *Social Forces*, 1940, 18, 560–568.

MINER, H. The folk-urban continuum. *American Sociological Review*, 1952, 17, 529–537.

PARSONS, T., & SHILS, E. A. (Eds.) *Toward a general theory of action*. Cambridge: Harvard University Press, 1959.

REDFIELD, R. *The folk culture of Yucatan.* Chicago: University of Chicago Press, 1941.

REDFIELD, R. The folk society. *American Journal of Sociology,* 1947, **52**, 293–308.

RIESMAN, D. *The lonely crowd.* New Haven: Yale University Press, 1950.

STOUFFER, S. A., & TOBY, J. Role conflict and personality. *American Journal of Sociology,* 1951, **56**, 395–406.

WHITEHILL, A. M., JR. Cultural values and employee attitudes: United States and Japan. *Journal of Applied Psychology,* 1964, **48**, 69–72.

WHYTE, W. F. Culture, industrial relations, and economic development: The case of Peru. *Industrial and Labor Relations Review,* 1963, **16**, 583–594.

WHYTE, W. F., & WILLIAMS, L. K. Supervisory leadership: An international comparison. Symposium B3, Paper No. B3C, New York State School of Industrial and Labor Relations, Cornell University, 1963, 1–8.

WHYTE, W. H. *The organization man.* New York: Harper, 1948.

ZURCHER, L. A., MEADOW, A., & ZURCHER, S. L. Value orientation, role conflict, and alienation from work: A cross-cultural study. *American Sociological Review,* 1965, **30**, 539–548.

(Received April 3, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 2, 145–147

ACHIEVEMENT VALUES AND SCIENTIFIC PRODUCTIVITY

BISHWA NATH MUKHERJEE [1]

Nagpur University, Nagpur, India

50 college science teachers were tested on Mukherjee's Sentence Completion Test (SCT), a forced-choice scale for measuring achievement values (v Ach). The hypothesis of a positive relationship between v Ach and scientific productivity was tested. Results indicated a significant positive correlation between the 2 variables. Most of the teachers having advanced research degrees were also found to have high v Ach scores. The study provides additional evidence of the construct validity of the SCT as a measure of v Ach and demonstrates the importance of v Ach in the prediction of scientific productivity.

The purpose of the present study is to investigate the relationship between achievement values (v Ach) of university teachers in the science faculty and their productivity in written scientific work defined in terms of the number of publications in scientific journals, research reports, technical memos, and papers contributed to scientific meetings and conferences (Taylor, Smith, & Ghiselin, 1963). The research is based on the premise that achievement values can best be viewed as those principles that guide one's conduct in making substantive contribution to the field or profession in which one is engaged. It was therefore predicted that achievement values of science teachers and the number of their scientific publications are positively correlated. In addition to verifying the above hypothesis, the present study aimed at exploring the relationship, if any, between achievement values and academic degrees earned.

METHOD

Subjects

Fifty male teachers of various science departments of Patna University, Patna, India volunteered as *S*s for the study. Their ages ranged from 22 to 55 yr. with a mean of 31.86 yr. Out of the 50 teachers who were sampled for the study, 28 teachers were found to participate in both undergraduate and graduate programs. Of the remaining, 17 were primarily engaged in teaching at the undergraduate level and 5 were exclusively connected with graduate programs.

[1] The author wishes to convey his heartfelt thanks to Braj Kumar Mishra, lecturer in psychology, Sindri College, Sindri, India for collecting the data reported in this article.

Measurement of Achievement Values

Achievement values were measured by the Sentence Completion Test (SCT), a forced-choice scale, developed by Mukherjee (1964). The SCT consists of 50 forced-choice triads (one item reflecting achievement-related sentiment and the other two pertaining to different aspects of manifest needs) selected in such a manner as to minimize the "social desirability" factor. A few examples of the types of forced-choice triads constituting the SCT are given below:

1. I like .....
   - A. to be faithful to my friends and colleagues.
   - B. to be very systematic in my work.
   - C. to do my best in whatever work I undertake.
2. I avoid .....
   - A. gay and irresponsible pleasure-seekers.
   - B. mental confusion and chaos.
   - C. those situations which are not competitive.
3. I have a general tendency .....
   - A. to continue a work till it is finished.
   - B. to analyze others' judgments critically.
   - C. to be polished in my manners.

The *S*'s score on SCT is the number of times he has chosen a statement reflecting achievement-related values. Details concerning the construction, administration, scoring, and various psychometric properties of the SCT have been reported by Mukherjee (1964, 1965c). The test-retest reliability of SCT after an interval of nearly 2 mo. has been found in various studies to range from .71 to .83. The test-retest correlation has been found to be almost .75 for a sample of 100 Indian male graduates to whom the test was readministered after a month and a half (Mukherjee & Sinha, 1967). The Kuder-Richardson reliability estimate for the SCT

was .716 for a sample of 248 American college students enrolled at the Jeffersonville and Bloomington Campus of Indiana University. Previous studies (Mukherjee, 1964, 1965a, 1965c, 1965d, 1966) have demonstrated the concurrent and construct validities of SCT as a measure of v Ach. A factor analysis of SCT (Mukherjee, 1965b) revealed that the test covers some of the important dimensions of cognitive expectations regarding achievement, such as, aspiration, ego-ideal, perseverance, identification with a successful authority, preference for difficult tasks, etc.

Procedure

Mimeographed copies of SCT along with separate answer sheets were distributed randomly to 50 science teachers regardless of their age and fields of specialization. They were requested to complete the SCT during their leisure time and return the answer sheet within a couple of days. All answer sheets were turned in within 2 days. On the answer sheet, each *S* provided information regarding his (*a*) present academic rank, (*b*) length of service as a college teacher, (*c*) academic degrees, (*d*) level of teaching, (*e*) age, and (*f*) publications. Papers published in scientific journals only defined each teacher's level of scientific productivity. Papers contributed to various scientific meetings were also considered, irrespective of whether they were published or not.

RESULTS

Based on the total scores on SCT, *S*s were classified into two criterion groups, High v Ach and Low v Ach. The 25 *S*s scoring 22 and above were arbitrarily designated as the High group and the remaining 25 *S*s scoring 21 and below formed the Low v Ach group. Table 1 presents the means for the two groups on some of the selected variables.

The data on number of scientific publications were examined by *t* test (one-tailed) for differences between means and by correlational analysis. However, the Student *t* test could not be applied directly since the High v Ach group had a significantly higher variance in publication variable. So in order to achieve homoscedasticity, the number of papers (Y_i) contributed by each *S* was subjected to a transformation of the form: log ($Y_i + 1$). The means of these transformed scores for the High and Low v Ach groups were .7001 and .3665, respectively. The unbiased variances turned out to be .2541 and .2478, respectively. A *t* test indicated that the difference between the two criterion groups with respect to transformed scores on the publication variable was statistically significant ($t = 2.36$, $p < .02$). Computation of the product-moment correlation between SCT scores and number of publications also indicated a significant positive relationship ($r = .32$, $p < .02$).

In order to test the hypothesis of no association between v Ach as measured by SCT and degrees earned, a chi-square test was run. The χ^2 turned out to be 4.024 which for 1 *df* is significant beyond .05 level. When the data on level of teaching were subjected to a chi-square analysis, it was found that most of the teachers belonging to the High v Ach group participated in the graduate program whereas most *S*s having low v Ach scores were exclusively engaged in undergraduate teaching ($\chi_1^2 = 7.20$, $p < .01$). The hypothesis of no association between v Ach and academic rank could not be rejected ($\chi_1^2 = 2.10$, $p > .10$). Similarly, age and v Ach scores were found to have no relationship ($\chi_1^2 = .32$). A chi-square test also revealed that the two criterion groups did not differ significantly with respect to length of teaching experience ($\chi_1^2 = .72$).

TABLE 1

MEANS FOR THE TWO CRITERION GROUPS ON A FEW SELECTED VARIABLES

Variable	High v Ach group (N=25)	Low v Ach group (N=25)
SCT score	23.96	16.80
No. publications	8.20	3.80
Teaching experience in years	8.70	9.49
Age in years	31.00	32.72

DISCUSSION

The study demonstrates that college teachers in science having high achievement values show greater scientific productivity than their colleagues whose v Ach scores are low. Conversely, the more productive *S*s of the present study have attitudes which suggest that they strive for a long record of professional achievement, for more distant goals, and a stronger desire to compete successfully with a standard of excellence than the less pro-

ductive *Ss* even though the opportunities for scientific research are comparable for both the groups. Since the above relationship was predicted in advance, the study provides additional evidence of the construct validity of the SCT as a measure of achievement values.

Direct validational evidence is also furnished by some of the other data reported here. That the relationship between SCT scores and academic degrees earned should be positive is quite consistent with the concept of achievement values according to which a person having high achievement orientation is one who shows self-actualizing tendency (Maslow, 1954), exercises his talents fully and enthusiastically to achieve higher goals that are related to his further self-development. It is no wonder that teachers having higher scores on the SCT were found more frequently to teach at the graduate level than those who have low scores on the SCT. The finding holds true even when factors such as age, teaching experience, and academic rank are held constant. The result is interpreted to mean that *Ss* of the High v Ach group have a higher level of aspiration and a keen interest in undertaking difficult tasks. The study lends conclusive evidence for the hypothesis that the productive scientist has a higher achievement value, and thus individual differences with regard to scientific productivity may be accounted for not only in terms of ability and opportunity but also in terms of their achievement values.

REFERENCES

MASLOW, A. H. *Motivation and personality*. New York: Harper, 1954.

MUKHERJEE, B. N. Development of a forced-choice test of achievement motivation. Final report, United States Office of Education, Washington, D. C. Contract No. OE-4-10-233, Project S-113, Indiana University, Bloomington, August 1964.

MUKHERJEE, B. N. Achievement motivation and goal-setting behavior in the classroom. *British Journal of Educational Psychology*, 1965, **35**, 286–293. (a)

MUKHERJEE, B. N. Factorial analysis of a forced-choice test of achievement motivation. *Journal of Vocational and Educational Guidance*, 1965, **11**, 43–54. (b)

MUKHERJEE, B. N. A forced-choice test of achievement motivation. *Journal of Indian Academy of Applied Psychology*, 1965, **2**, 85–92. (c)

MUKHERJEE, B. N. Self perception and achievement motivation. *Journal of Psychological Researches*, 1965, **9**, 109–114. (d)

MUKHERJEE, B. N. Learning and speed of performance as a function of need for achievement. *Journal of Indian Academy of Applied Psychology*, 1966, **3**, 23–27.

MUKHERJEE, B. N., & SINHA, B. N. Some relationships between anxiety level and need for achievement. *Journal of Indian Academy of Applied Psychology*, 1967, **4**, 11–21.

TAYLOR, C. W., SMITH, W. R., & GHISELIN, B. The creative and other contributions of one sample of research scientists. In C. W. Taylor & F. Barron (Eds.), *Scientific creativity: Its recognition and development*. New York: Wiley, 1963. Pp. 53–76.

(Received April 11, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 2, 148-153

DIFFERENTIAL EFFECT OF NOISE ON TASKS OF VARYING COMPLEXITY [1]

DAVID H. BOGGS [2] AND J. RICHARD SIMON

University of Iowa

The method of simultaneous tasks was used to test the hypothesis that the deleterious effect of noise on performance varies as a function of task complexity. 48 Ss performed on 1 of 2 complexity levels of a 4-choice RT task and, at the same time, performed a secondary auditory monitoring task. All Ss performed in both quiet and in noise. Performance indexes were RT and secondary-task errors. Noise produced a significantly greater increase in secondary-task errors when the secondary task was paired with the complex primary task than when it was paired with the simple primary task. Secondary-task performance provided a more sensitive measure of both task complexity and the effect of noise than did the RT measure.

The purpose of this experiment was to test the hypothesis that the deleterious effect of noise on performance increases as a function of increasing task complexity. There have been numerous studies conducted to investigate the effect of environmental noise on performance. Some have indicated a decremental effect of noise (e.g., Broadbent, 1953, 1954, 1958a; Jerison, 1959; Jerison & Wing, 1957; Sanders, 1961; Woodhead, 1964), while others have indicated an incremental effect (e.g., Kirk & Hecht, 1963; Wilbanks, Webb, & Tolhurst, 1956). A substantial number of studies, however, have revealed no significant effect whatsoever (e.g., Broadbent, 1957; Hyman, 1950; Jerison, 1957; Kryter, 1950; Kurz, 1964; Wilbanks, Webb, & Tolhurst, 1956). If research which has remained unpublished as a consequence of achieving negative results were to be made available (e.g., Chisman, 1963), the weight of evidence indicating that noise does not significantly affect performance would be considerably increased.

One explanation for the prevalence of negative results in laboratory studies involving noise might be that *S*s were able to draw upon unused perceptual capacity and thereby maintain their performance despite the distracting noise. In order to detect changes in unused capacity associated with performance in noise, the present study utilized a procedure called the "method of simultaneous tasks." The procedure involved having *S* perform, simultaneously with the primary experimental task, a secondary task which effectively occupied that portion of *S*'s perceptual capacity which was not occupied by the requirements of the primary task. Thus, performance on the secondary task provided a measure of unused capacity, and any effect of noise sufficient to require the use of this excess capacity to maintain primary-task performance was evidenced as a decrement in secondary-task performance. According to Poulton (1965), this method was first employed by Bornemann (1942a, 1942b) although it was apparently also developed independently by Bahrick, Noble, and Fitts (1954).

Of special relevance to the present research are two studies which utilized the method of simultaneous tasks. Poulton (1958) used secondary-task performance as a measure of the amount of attention required by a perceptual-motor task and was able to demonstrate a difference in difficulty between two versions of the task which was not apparent by comparing performance on the two tasks when performed alone. Broadbent (1964) employed the method to investigate the relationship between attention required to a reaction-time (RT) task and degree of stimulus-response (S-R) compatibility. Using a secondary, mental-arithmetic task, he compared the performance of *S*s working on compatible and incompatible versions of an RT

[1] This paper is based upon a thesis completed by the first author under the direction of the second author and submitted to the University of Iowa in partial fulfillment of the requirements of the master's degree.

[2] Now at the University of Connecticut.

task. The incompatible task caused more interference with the secondary task than did the compatible task demonstrating that a perceptual-motor task may be made more difficult by making the S-R code less compatible, and that this difference in difficulty may be measured indirectly by observing performance on a simultaneous, secondary task.

There are probably many reasons for the apparent conflicting results concerning the effects of noise on performance. One important factor seems to be that some tasks are more vulnerable to noise than others. According to Broadbent (1958b), noise produces brief failures of perception which, in turn, result in the marked decrements observed on vigilance tasks and tasks which require continuous handling of information. Jerison (1957) suggests that the primary factor determining the effect of noise on performance may be the "flexibility of attention" required by the task. Noise seems to have a decremental effect on vigilance tasks which involve several sources of information whereas tasks which require attention to only one information source are not affected.

Task difficulty may also be important in determining whether noise has a decremental effect on performance. Broadbent (1954) found a noise-induced decrement on a "difficult" watch-keeping task (Twenty Dials Test) but not on an "easy" task (Twenty Lights Test). Park and Payne (1963), however, found no evidence of an interaction between noise conditions and the level of difficulty of division problems.

The present experiment was concerned with the variable of difficulty or, more properly, task complexity. In this study, *Ss* performed two tasks simultaneously. The overall complexity or perceptual load imposed by these simultaneous tasks was manipulated by altering the S-R compatibility of the primary task, a four-choice RT task. Performance on a secondary, auditory-monitoring task was used to measure *Ss*' available excess perceptual capacity. It was hypothesized that the introduction of noise would increase the total perceptual load, reduce *Ss*' reserve capacity, and thereby result in a performance decrement on the secondary task. It was further hypothesized that this noise-induced performance decrement on the secondary task would be more marked under the condition of high task complexity.

METHOD

All *Ss* performed two tasks simultaneously under quiet conditions and also while exposed to an intermittent, annoying noise. The primary task was of the perceptual-motor type and was presented in either of two levels of complexity. The secondary auditory-monitoring task was held constant under these two complexity levels and was used to detect changes in overall perceptual load.

Primary Task

The primary task was a four-choice reaction-time (RT) task in which *Ss*, after being alerted by a green warning light, responded as rapidly as possible to the onset of one of four red stimulus lights by pressing the appropriate one of four switches which were located directly below each of the lights. In the simple condition, the correct switch corresponded to the light directly above it. In the complex condition, there was no spatial correspondence between light and switch. In this latter condition, *Ss* were instructed to respond: (1) to the left light by pressing the second switch from the left; (2) to the second light from the left by pressing the switch at the extreme left; (3) to the second light from the right by pressing the extreme right switch; and (4) to the right light by pressing the second switch from the right. The four stimulus lights were arranged 3½ in. apart in a horizontal row, located at a distance of approximately 24 in. from *S*'s eyes, slightly below shoulder level.

Presentation of warning and stimulus lights in the RT task was controlled automatically by three Hunter interval timers. The first timer activated the warning light for a duration of 0.5 sec. The second timer provided a 1.0-sec. interval between warning light and stimulus light and also activated the stimulus light. The third timer provided a 5.0-sec. delay following the onset of the stimulus light, and then automatically initiated a recycling of the three timers. A Hunter Klockounter began at the onset of the stimulus light and stopped when *S* pressed one of the switches, thus providing a measure of RT. Pressing the switch also extinguished the stimulus light. The order of presentation of stimulus lights was controlled manually by *E* according to a predetermined sequence. The lights were presented in this same sequence on every block of trials, each of the four stimulus lights appearing 23 times. Signal lights on *E*'s side of the apparatus permitted detection of errors in *S*'s choice of switch. Since pressing any of the four switches caused the stimulus light to be extinguished, *S* received no information concerning the correctness of his responses.

The *S* pressed the switches with the index finger of his preferred hand. To control hand-travel time, *S* began each trial from a rest position which was marked on the table 7 in. in front of the RT

apparatus, on an imaginary line bisecting, and perpendicular to, a line passing through the four switch knobs.

Secondary Task

The secondary task was similar to one used by Brown (1965). A continuous sequence of digits was presented to *S* through a headphone placed over his left ear. He was instructed to listen for odd-even-odd sequences of digits, and to respond immediately by saying "yes" whenever such a sequence was detected. The digits were recorded on one channel of a stereophonic magnetic tape at regular intervals of 1.5 sec. One odd-even-odd sequence was located randomly within every 15-sec. period of the 10-min. tape, so that the tape contained a total of 40 odd-even-odd sequences. The second channel contained signals, audible only to *E*, indicating the location of each odd-even-odd sequence. The tape was played on a Sony Model 464 stereophonic recorder. The *S* listened to the digits channel through a Grason-Stadler Model TDH-39 earphone (equipped with MX 41-AR rubber cushion) connected directly to the output of the appropriate internal playback amplifier of the recorder, while *E* listened to the marker signals through headphones connected through a Sony external power amplifier to the playback-preamplifier output of the second channel. The playback-preamplifier gain of the channel containing the signals was reduced to a minimum to prevent leakage into the digits channel. This arrangement permitted constant synchronization of the digits with the signals to *E*, and provided complete consistency throughout the experiment in the scoring of errors on the secondary task.

Noise

The noise consisted of recorded 0.5-sec. bursts of the sound produced by a bandsaw cutting aluminum. Acoustical frequency analysis revealed strong peaks at approximately 1800 and 1900 cps. A total of 90 bursts (92 db. SPL) were presented in the course of each 10-min. experimental run. These bursts were recorded on the tape containing the digits. The bursts were located randomly, but with the restriction that they never overlap any of the secondary-task digits. The masking effect of noise, therefore, was not a factor in this experiment. The digits on the "noise" tape were copied directly from the "quiet" tape, using a second recorder, so that the only difference between the two tapes was, in fact, the presence of noise on the "noise" tape.

Subjects

The *S*s were 24 males and 24 females from the elementary psychology courses at the University of Iowa. All *S*s reported having normal hearing.

Experimental Design

A mixed design involving one within- and three between-*S* dimensions was employed (Lindquist, 1953). Twelve men and 12 women performed on the simple RT task, and 12 men and 12 women performed on the complex RT task. Within each complexity condition, 12 *S*s (6 men and 6 women) performed first in quiet and then in noise, while the remaining 12 *S*s (6 men and 6 women) performed first in noise and then in quiet. Complexity level and sequence were assigned randomly to *S*s within each sex group.

Procedure

Each *S* served in one experimental session consisting of one run in quiet and one in noise. Each run was 10 min. in duration. The *S* listened first to a 1-min. tape demonstrating the correct method of responding to the digits task. He was then given two consecutive 1-min. practice sessions on the digits task alone. If he failed to respond without error during the first min. of practice, the tape was stopped and the instructions repeated. All *S*s succeeded in performing without error during the second min. of practice. Next, the RT-task instructions were read, followed by the appropriate instructions for the first experimental condition, that is, quiet or noise. In order to establish the relative primacy of the RT task, *S* was instructed repeatedly that he should do as well as he could on the digits task, but that he should not let the digits task interfere with the RT task. After completing the first run, the instructions for the second condition were read, and *S* was again cautioned not to let the digits task interfere with the RT task. The entire session lasted approximately 45 min.

Scoring

Records were kept of RT, errors on the RT task, and errors on the digits task. Omissive errors (failure to report the occurrence of an odd-even-odd sequence in the 1.5-sec. interval after the completion of the sequence) and commissive errors (inappropriate "yes" responses) on the digits task were recorded separately. The first 12 RT trials of both experimental runs were considered practice trials and were not included in the calculation of RT medians and error totals.

RESULTS

Performance data from both primary and secondary tasks were analyzed to determine the effects of noise on performance.

Primary Task

An analysis of variance [3] indicated no significant effect of noise on RT (787 vs. 784

[3] Summaries of analyses of variance and other summary tables have been deposited with the American Documentation Institute. Order Document No. 9747 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washing-

TABLE 1

TOTAL ERRORS ON SECONDARY TASK AS A FUNCTION OF SEQUENCE, NOISE, TASK COMPLEXITY, AND SEX

Task	Sequence	Males			Females			Overall *M*
		Condition			Condition			
		Quiet	Noise	*M*	Quiet	Noise	*M*	
Simple	Quiet-Noise	7.67	11.83	9.75	10.33	13.67	12.00	10.88
	Noise-Quiet	8.67	11.17	9.92	7.33	10.33	8.83	9.38
	M	8.17	11.50	9.83	8.83	12.00	10.42	10.13
Complex	Quiet-Noise	14.17	21.00	17.58	11.67	18.33	15.00	16.29
	Noise-Quiet	8.83	17.83	13.33	10.33	17.00	13.67	13.50
	M	11.50	19.42	15.46	11.00	17.67	14.33	14.90
Combined mean for both complexity levels		9.83	15.46	12.65	9.92	14.83	12.38	

msec.). There was a tendency for males to respond faster than females (766 vs. 806 msec.), and for RT on the complex task to be slower than on the simple task (808 vs. 763 msec.), but neither of these effects was significant.[4] The only significant factor affecting RT was order; $F(1/40) = 11.76$, $p < .01$, with the first run being slower than the second (805 vs. 766 msec.). No errors were made on the primary task during the experimental runs.

Secondary Task

Table 1 shows the effects of sequence, noise, task complexity, and sex on the total number of secondary-task errors, that is, omissive plus commissive errors. It may be seen that mean error totals for all sex-sequence groups were higher in noise than in quiet, and that, in general, a greater number of errors was made under the complex task condition than under the simple condition. An analysis of variance revealed that the main effect of noise was significant; $F(1/40) = 62.52$, $p < .001$. The analysis also indicated a significant main effect of complexity; $F(1/40) = 8.31$, $p < .01$. Of major interest, however, was a significant Noise × Complexity interaction; $F(1/40) = 9.19$, $p < .01$. Table 2 summarizes this effect. It will be noted that the introduction of noise produced a greater increase in errors under the complex task condition than under the simple task condition. There was no sequence effect to complicate the interpretation.

Separate analyses of variance were performed for omissive and for commissive errors on the secondary task. Results of the analysis of omissive errors were identical to those of the total errors; that is, the effects of both noise and task complexity were sig-

TABLE 2

EFFECT OF NOISE ON SECONDARY-TASK PERFORMANCE UNDER SIMPLE AND COMPLEX TASK CONDITIONS

Task	Mean errors			Difference (noise minus quiet)
	Quiet	Noise	Overall *M*	
Simple	8.50	11.75	10.13	3.25
Complex	11.25	18.54	14.90	7.29

ton, D. C. 20540. Remit in advance $1.25 for microfilm or $1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

[4] In a separate experiment, 20 *S*s performed the RT task alone, that is, without the simultaneous secondary task. Half of the *S*s performed the simple version of the task while the other half performed the complex version. Performance on the complex task was significantly slower than on the simple task (708 vs. 587 msec.; $t = 2.45$, $p < .05$).

nificant, and there was a significant Noise × Complexity interaction. In the analysis of commissive errors, noise was the only significant factor.

DISCUSSION

Results of this study demonstrate that noise produces a significant decrement in performance and that the magnitude of this decrement varies as a function of task complexity. Results also suggest that noise effects may have gone undetected in much previous research for lack of a sensitive criterion measure.

Many tasks demand less than the total perceptual capacity which an individual has available. In these situations, the person possesses what may be termed a reserve capacity, which is free from the demands of the task. Should the perceptual demands of the task or of the situation increase, he may draw upon this reserve capacity and, by so doing, may be able to maintain task performance at a constant high level. In the present study, the fact that increased task complexity, and the introduction of noise, did not affect RT performance is attributed to this compensatory mechanism.

The method of simultaneous tasks provides a means of detecting changes in reserve capacity associated with changes in the task and/or environment. Secondary-task performance was used in this study as a measure of reserve perceptual capacity. In this experiment, the complex primary task made greater perceptual demands than the simple primary task. Therefore, less reserve capacity was available for the secondary task when it was paired with the complex task than when it was paired with the simple task. This resulted in the greater number of errors on the secondary task when it was paired with the complex primary task. The introduction of noise "used up" some of *S*'s reserve capacity; that is, *S* had to draw from his reserve so that primary-task performance would not suffer as a consequence of the noise. There was, therefore, less perceptual capacity available for secondary-task performance, and, consequently, errors increased. The fact that the noise produced a greater increase in secondary-task errors when the secondary task was paired with the complex primary task than when it was paired with the simple primary task may be attributed to the fact that the noise occupied proportionally more of the reserve capacity in the former case than in the latter.

REFERENCES

BAHRICK, H. P., NOBLE, M., & FITTS, P. M. Extra-task performance as a measure of learning a primary task. *Journal of Experimental Psychology,* 1954, **48**, 298–302.

BORNEMANN, E. Untersuchungen über den Grad der geistigen Beanspruchung: Ausarbeitung der Methode. *Arbeitsphysiologie,* 1942, **12**, 142–172. (a)

BORNEMANN, E. Untersuchungen über den Grad der geistigen Beanspruchung: praktische Ergebnisse. *Arbeitsphysiologie,* 1942, **12**, 173–191. (b)

BROADBENT, D. E. Noise, paced performance and vigilance tasks. *British Journal of Psychology,* 1953, **44**, 295–303.

BROADBENT, D. E. Some effects of noise on visual performance. *Quarterly Journal of Experimental Psychology,* 1954, **6**, 1–5.

BROADBENT, D. E. Effects of noise on behavior. In C. M. Harris (Ed.), *Handbook of noise control.* New York: McGraw-Hill, 1957. Ch. 10.

BROADBENT, D. E. Effect of noise on an "intellectual" task. *Journal of the Acoustical Society of America,* 1958, **30**, 824–827. (a)

BROADBENT, D. E. *Perception and communication.* New York: Pergamon Press, 1958. (b)

BROADBENT, D. E. S-R compatibility and the processing of information. *Acta Psychologica,* 1964, **23**, 325–327.

BROWN, I. D. A comparison of two subsidiary tasks used to measure fatigue in car drivers. *Ergonomics,* 1965, **8**, 467–473.

CHISMAN, J. A. The effects of continuous noise on the performance time and forces exerted in performing a subminiature assembly task. Unpublished doctoral dissertation, University of Iowa, Iowa City, 1963.

HYMAN, D. E. An investigation of the effect of various noise levels on performance on three mental tasks. Unpublished master's thesis, Purdue University, Lafayette, Indiana, 1950. Cited by E. J. McCormick, *Human factors engineering.* New York: McGraw-Hill, 1964. P. 475.

JERISON, H. J. Performance on a simple vigilance task in noise and quiet. *Journal of the Acoustical Society of America,* 1957, **29**, 1163–1165.

JERISON, H. J. Effects of noise on human performance. *Journal of Applied Psychology,* 1959, **43**, 96–101.

JERISON, H. J., & WING, S. Effects of noise on a complex vigilance task. USAF WADC Tech. Rep. TR 57-14, AD 110700, 1957.

KIRK, R. E., & HECHT, E. Maintenance of vigilance by programmed noise. *Perceptual and Motor Skills,* 1963, **16**, 553–560.

KRYTER, K. D. The effects of noise on man. *Journal of Speech and Hearing Disorders Monograph Suppl. 1,* 1950, 1–95.

KURZ, R. B. Effects of three kinds of stressors on human learning and performance. *Psychological Reports,* 1964, 14, 161–162.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

PARK, J. F., & PAYNE, M. C., JR. Effects of noise level and difficulty of task in performing division. *Journal of Applied Psychology,* 1963, 47, 367–368.

POULTON, E. C. Measuring the order of difficulty of visual-motor tasks. *Ergonomics,* 1958, 1, 234–239.

POULTON, E. C. On increasing the sensitivity of measures of performance. *Ergonomics,* 1965, 8, 69–76.

SANDERS, A. F. The influence of noise on two discrimination tasks. *Ergonomics,* 1961, 4, 253–258.

WILBANKS, W. A., WEBB, W. B., & TOLHURST, G. C. A study of intellectual activity in a noise environment. United States Naval School of Aviation Medicine, Research Project NM 001 104 100, Report 1, October 31, 1956.

WOODHEAD, M. M. The effect of bursts of noise on an arithmetic task. *American Journal of Psychology,* 1964, 77, 627–633.

(Received April 17, 1967)

Manuscripts Accepted for Publication in the *Journal of Applied Psychology*

Individual versus Group Problem Solving in an Industrial Sample: John P. Campbell*: College of Liberal Arts, Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455.

Comparison of Judgmental Methods with Endorsements in the Assessment of Personality Traits: Igor Kusyszyn*: 135A Behavioural Sciences Building, York University, 4700 Keele Street, Downsview, Ontario.

Need Type and Job Satisfaction among Industrial Research Scientists: George B. Graen*, Rene V. Dawis, and David J. Weiss: Department of Psychology, University of Illinois, Urbana, Illinois 61801.

Age, Cigarette Smoking, and Tests of Physical Fitness: Kenneth H. David*: University of Maryland, Far East Division, APO San Francisco 96525.

Ability, Values, and Concepts of Equitable Salary Increases in Exercise Compensation: Bernard M. Bass*: 401 Bruce Hall, Graduate School of Business, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

Evaluation of a New Hybrid in Programmed Instruction: D. H. Schuster*: Department of Psychology, Iow State University, Ames, Iowa 50010.

Absolute Scaling of Job Performance: Arthur I. Siegel*, Douglas G. Schultz, M. A. Fischl, and Richard S. Lanterman: Applied Psychological Services, Wayne, Pennsylvania.

Comparison of Operator Performance on a Tracking Task Using Discretely and Continuously Variable Controls: M. Hammerton* and A. H. Tickner: Medical Research Council, Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.

Vehicle Kinematics and Gap Acceptance: Paul M. Hurst*, Kenneth Perchonok, and Edmond L. Sequin: Institute for Research, 257 South Pugh Street, State College, Pennsylvania 16801.

Searching for Letters or Closed Shapes in Simulated Electronic Displays: E. C. Poulton*: Assistant Director, Medical Research Council, Applied Psychology Research Unit, Cambridge, England.

Addendum to "An Empirical Investigation of Two Implications of the Two-Factor Theory of Job Satisfaction": George B. Graen* and Charles L. Hulin: Department of Psychology, University of Illinois, Urbana, Illinois 61801.

Testing Traditional and Two-Factor Hypotheses Concerning Job Satisfaction: George B. Graen*: Department of Psychology, University of Illinois, Urbana, Illinois 61801.

Signs, Samples, and Criteria: Paul F. Wernimont and John P. Campbell*: College of Liberal Arts, Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455.

Effect of Defect Rate on Inspection Accuracy: Douglas H. Harris: Bill A. Wheeler, Assistant Director, Public Relations, Autonetics Division, 3370 Miraloma Avenue, Anaheim, California 92803.

---

* Asterisk indicates author for whom the address is supplied.

Journal of Applied Psychology
1968, Vol. 52, No. 2, 154–157

EXPECTATIONS AND DROPOUTS IN SCHOOLS OF NURSING

MILDRED ENGBERG KATZELL [1]

National League for Nursing, New York City

The expectations and experiences of stress and satisfaction of 1,852 1st-yr. students in 43 schools of nursing were assessed by questionnaire. Low but significant negative correlations were obtained between withdrawal and (a) experienced satisfactions and (b) confirmation of expectations, especially with respect to satisfactions. Withdrawal was unrelated to experienced stress, whether expected or unexpected.

Remarks about the nursing shortage are by now commonplace. Since 20–25% of entrants to schools of nursing withdraw during the first year, reducing this attrition rate would alleviate the shortage.

Fox, Diamond, and Walsh (1960) reported on the situations which student nurses had indicated as stressful and satisfying, including those elements in the school situation to which students alleged their withdrawal was due. In this report and elsewhere, students who disliked nursing indicated that it was not what they had expected.

In addition, Kibrick (1958) observed that the satisfactions expected by nursing students did not have much bearing on their remaining in school, but that those who were more realistic in their appraisal of nursing were more likely to remain. This finding is consistent with that of Weitz (1956) who discovered that life insurance agents were more likely to remain in their jobs if they had a clear picture of the job duties. Block, Yuker, Campbell, and Melvin (1964) and Brown and Landsberger (1960) found that dissatisfaction is influential in determining whether an individual will withdraw during the early stages of employment. Block and his colleagues added that the longer the worker remains with the company, the less important satisfaction becomes in determining his further continuation. This may also be true of student nurses, among whom the withdrawal rate is greatest during the first year, and declines markedly thereafter.

Among the theories that have dealt with withdrawal from group participation are those of Ross and Zander (1957), March and Simon (1958), Herzberg, Mausner, and Snyderman (1959), Katzell (1964), and Vroom (1964).

Degree of need satisfaction was hypothesized by Ross and Zander to be related to continued participation in the employment situation. Herzberg et al., on the other hand, classified the elements of the work situation into satisfiers or motivating factors, and dissatisfiers or hygiene factors. Turnover, in this theory, was viewed as being related to a low level of hygiene factors, causing dissatisfaction, whereas the motivating factors, or satisfiers, were perceived as being related primarily to productivity.

In Katzell's view, withdrawal is directly related to the experienced or expected dissatisfactions associated with the situation and the satisfactions associated with an alternative situation. He also theorized that withdrawal is negatively related to the experienced or expected satisfactions associated with the situation and to the dissatisfactions associated with an alternative situation. Vroom postulated that the more satisfied the individual, the greater the force on him to remain in the situation and the less the probability of his voluntarily withdrawing from it. He also indicated that decisions to remain or withdraw are related to the positive and negative valences of outcomes and to the expectancies that the outcomes will be attained.

March and Simon specified that continued participation depends upon the difference between rewards and contributions, and sug-

[1] This article is based upon data collected by the author in fulfilling the dissertation requirements for the PhD degree at Teachers College, Columbia University.

gested that the difference be estimated from measures of satisfaction, since the greater the excess of rewards over contributions, the greater should be satisfaction. Satisfaction was seen as being related to, among other factors, predictability in terms of the degree to which expectations are realized.

OBJECTIVES

This study was undertaken as a partial test of the March and Simon theory, as it applies to withdrawal from schools of nursing. From the theory, it should follow that the greater the number of satisfactions experienced, the more likely a student will be to continue in a school of nursing. Conversely, the greater the number of stresses experienced, the more likely a student will be to withdraw. In addition, a student will be more likely to withdraw if she experiences a large number of unexpected stresses than if she experiences the same number of stresses but they were expected. It should also be true that expected but unrealized satisfactions will tend to cause withdrawal, and that experiencing expected satisfactions will do more to prevent withdrawal than experiencing unexpected satisfactions.

METHOD

Sample

The sample comprised 1,852 first-year students in 43 schools of nursing. Among the 1,852 students, 1,439 (77.7%) returned for the second year; 183 (9.9%) withdrew for academic reasons; 230 (12.4%) withdrew for nonacademic reasons.

Instruments

A 139-item questionnaire was constructed to measure the satisfactions (56 items) and stresses (83 items) expected and experienced by the students. A single set of items was used, with one form of instructions being given when the students' expectations were collected, and another when experiences were collected.

Illustrative of the questionnaire items are the following 10, of which the first 5 were classified as stress items, and the remainder as satisfaction items.

1. Being homesick.
2. Being respected less as a student nurse, by people outside the school, than if one had pursued another career.
3. Spending very little time working with patients.
4. Doubting the wisdom of the choice of nursing as a career.
5. Being in the presence of death.
6. Having lots of fun.
7. Having someone to tell one's troubles to.
8. Finding that a student nurse is an important part of the team in the hospital.
9. Being so interested in nursing that one doesn't find it hard to learn what one needs to know.
10. Getting good grades.

Fourteen scores were derived for each student from the pair of questionnaires.[2]

Procedure

During the first week they were in school, the students completed the Expectations Questionnaire. A supply of Experiences Questionnaires was provided to each school of nursing to be administered to any students who withdrew between the group sessions of the Expectations and Experiences Questionnaires. Eight months after the Expectations Questionnaire, the Experiences Questionnaire was administered to students still enrolled. One year after the students had entered the schools of nursing, a questionnaire was sent to each cooperating school to learn which students had withdrawn and the reason for withdrawal of each.

The 14 scores derived from the questionnaires were correlated, using point-biserial coefficients, with (*a*) survival-withdrawal, (*b*) survival-academic withdrawal, and (*c*) survival-nonacademic withdrawal.

RESULTS

The scores of the 1,439 survivors and the three categories of dropouts are compared in Table 1. Differences significant at the .01 level were found between the mean scores of the survivors and each of the dropout groups in the total number of satisfactions experienced, the number of expected satisfactions experienced, and the number of expected satisfactions which were not experienced. In each case, those who remained in school had experienced more satisfactions. Although survivors and dropouts did not differ in numbers of expected satisfactions and stresses, the survivors had significantly more of their expectations confirmed than did the dropouts.

Survivors experienced more unexpected satisfactions than did dropouts, the difference being significant in the academic and total dropout groups. The academic dropouts experienced significantly more unexpected

[2] Copies of the questionnaires and data relative to their construction may be obtained from the author on request.

TABLE 1

SCORES OF 1,439 SURVIVORS, 413 DROPOUTS, 183 ACADEMIC DROPOUTS, AND 230 NONACADEMIC DROPOUTS

Score	A. Survivors		B. Dropouts		C. Academic dropouts		D. Non-academic dropouts		*t* ratios			
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	A–B	A–C	A–D	C–D
Satisfactions expected	37.68	6.64	37.33	6.99	37.99	6.89	36.81	7.01	.91	.57	1.76	1.71
Stresses expected	36.18	12.09	36.53	12.46	35.07	12.59	37.68	12.24	.51	1.13	1.72	2.11*
Satisfactions experienced	31.94	6.80	28.62	7.78	28.67	7.70	28.58	7.84	7.85**	5.47**	6.13**	.12
Stresses experienced	37.51	10.82	37.72	12.01	38.14	11.83	37.38	12.14	.32	.68	.15	.64
Expected satisfactions which were experienced	26.49	7.07	23.77	7.54	24.10	7.55	23.50	7.52	6.54**	4.05**	5.63**	.80
Expected satisfactions not experienced	11.19	4.92	13.57	6.07	13.89	6.26	13.31	5.90	7.30**	5.60**	5.16**	.96
Unexpected satisfactions which were experienced	5.45	3.66	4.85	3.79	4.57	3.23	5.08	4.16	2.86**	3.41**	1.27	1.40
Unexpected satisfactions not experienced	12.48	5.11	13.28	5.62	12.82	5.65	13.64	5.57	2.60**	.77	2.96**	1.47
Expected stresses which were experienced	24.13	9.66	24.04	10.40	23.57	10.27	24.42	10.48	.16	.70	.39	.83
Expected stresses not experienced	12.04	5.92	12.48	6.45	11.50	6.19	13.27	6.55	1.24	1.11	2.67**	2.81**
Unexpected stresses which were experienced	13.37	6.69	13.68	6.80	14.57	7.48	12.97	6.11	.82	2.06*	.91	2.33*
Unexpected stresses not experienced	33.04	11.09	32.08	11.79	32.52	12.04	31.73	11.59	1.48	.55	1.60	.67
Confirmed expectations	96.14	9.13	93.17	10.68	93.02	11.52	93.28	9.96	5.13**	3.51**	4.07**	.24
Unconfirmed expectations	42.04	9.02	44.58	10.19	44.52	10.82	44.63	9.66	4.53**	2.94**	3.77**	.10
Items omitted	.80	2.75	1.25	3.94	1.46	4.44	1.09	3.49	2.17*	1.96*	1.20	.92

* $p < .05$.
** $p < .01$.

stresses than either the survivors or the nonacademic dropouts, and expected significantly fewer stresses than the nonacademic dropouts.

The nonacademic and total dropout groups exceeded survivors in numbers of satisfactions which were neither expected nor experienced. Nonacademic dropouts also exceeded academic dropouts and survivors in the number of stresses expected but not experienced. The survivors omitted fewer items than the dropouts.

The point-biserial correlations between withdrawal and each of the 14 questionnaire

TABLE 2

CORRELATIONS BETWEEN WITHDRAWAL AND SCORES DERIVED FROM THE TWO QUESTIONNAIRES, BASED ON ALL SURVIVORS AND DROPOUTS

Score	All survivors and all dropouts (N = 1,852)	All survivors and academic dropouts (N = 1,622)	All survivors and nonacademic dropouts (N = 1,669)
Satisfactions expected	+.022	+.015	+.045
Stresses expected	+.012	−.029	+.043
Satisfactions experienced	−.193**	−.148**	−.164**
Stresses experienced	+.008	+.018	+.004
Expected satisfactions which were experienced	−.156**	−.105**	−.143**
Expected satisfactions not experienced	+.187**	+.165**	+.143**
Unexpected satisfactions which were experienced	−.067*	−.077*	−.034
Unexpected satisfactions not experienced	+.063*	+.021	+.077*
Expected stresses which were experienced	−.004	−.018	+.010
Expected stresses not experienced	+.030	+.029	+.070*
Unexpected stresses which were experienced	+.019	+.056	−.021
Unexpected stresses not experienced	−.035	−.014	−.040
Confirmed expectations	−.129**	−.104**	−.106**
Unconfirmed expectations	+.112**	+.084**	+.097**

* $p < .05$.
** $p < .01$.

scores are presented in Table 2. In each group, 5 of the 14 coefficients, though small, are significant at the .01 level. The 5 all pertain to satisfactions which were or were not experienced and to confirmation of expectations.

No relationship was demonstrated in any of the groups between withdrawal and (*a*) expected satisfactions, (*b*) expected stresses, (*c*) experienced stresses, (*d*) expected stresses which were experienced, (*e*) unexpected stresses which were experienced, and (*f*) unexpected stresses which were not experienced.

DISCUSSION

The results of this study supported March and Simon's (1958) postulates relative to satisfaction, but gave no support in relation to stress. The predicted effect of expectations was observed in the area of satisfactions, but received uncertain support in regard to stresses.

Contrary to Herzberg et al.'s (1959) view, it did not appear that student withdrawal was primarily due to dissatisfiers. It did, however, seem clear that there were differences between satisfiers and dissatisfiers, as Herzberg indicated, although no attempt was made here to validate Herzberg's measures.

The measures used in this study also did not lend themselves to direct implications for Ross and Zander's (1957) position. It would, however, be anticipated from their theory that dropouts, more than survivors, would report interference of school with their pursuit of extramural satisfactions, but this was not found to be the case.

The negative relationship between satisfaction and withdrawal, posited by Katzell (1964) and by Vroom (1964), received support from these findings. Their respective views concerning the relation of dissatisfaction, or negative valence, to withdrawal were not supported. The absence of stress did not have the favorable consequences associated with satisfaction, although the absence of satisfaction tended to have negative consequences, especially when the satisfaction was expected.

Despite the fact that significant correlations were obtained, so small a fraction of the total survival-withdrawal variance is predictable that the contribution may seem negligible. On the other hand, because of the large numbers of students and schools involved, the saving of even a small percentage of those who would otherwise withdraw from schools of nursing could contribute substantially in absolute numbers to the alleviation of the nursing shortage.

CONCLUSIONS

1. Withdrawal was inversely related to experienced satisfaction.
2. Withdrawal was inversely related to the confirmation of expectations, especially in the area of satisfactions.
3. Withdrawal was unrelated to experienced stresses, whether expected or unexpected.
4. Stress and satisfaction did not operate as opposite ends of the same continuum.

REFERENCES

BLOCK, J. R., YUKER, H. E., CAMPBELL, W. J., & MELVIN, K. B. Some correlates of job satisfaction among disabled workers. *Personnel and Guidance Journal*, 1964, **42**, 803–810.

BROWN, A. W., & LANDSBERGER, H. A. *The sense of responsibility among young workers.* (Reprint series No. 93) Ithaca: New York State School of Industrial and Labor Relations, Cornell University, 1960.

FOX, D. J., DIAMOND, L. K., & WALSH, R. C. *Satisfying and stressful situations in basic programs in nursing education.* New York: Institute of Research and Service in Nursing Education, Teachers College, Columbia University, 1960.

HERZBERG, F., MAUSNER, B., & SNYDERMAN, B. B. *The motivation to work.* New York: Wiley, 1959.

KATZELL, R. A. Personal values, job satisfaction, and job behavior. In H. Borow (Ed.), *Man in a world at work.* Boston: Houghton Mifflin, 1964. Pp. 341–363.

KIBRICK, A. K. Dropouts from schools of nursing: The effect of self and role perception. Unpublished doctoral dissertation, Harvard University, 1958.

MARCH, J. G., & SIMON, H. A. *Organizations.* New York: Wiley, 1958.

ROSS, I. E., & ZANDER, A. Need satisfactions and employee turnover. *Personnel Psychology*, 1957, **10**, 327–338.

VROOM, V. H. *Work and motivation.* New York: Wiley, 1964.

WEITZ, J. Job expectancy and survival. *Journal of Applied Psychology*, 1956, **40**, 245–247.

(Received April 28, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 2, 158–162

RELATION BETWEEN ETHNIC ORIGIN AND GSR REACTIVITY IN PSYCHOPHYSIOLOGICAL DETECTION [1]

SOL KUGELMASS AND ISRAEL LIEBLICH

Hebrew University of Jerusalem, Jerusalem, Israel

The psychophysiological data of 3 Jewish and 1 Bedouin samples were analyzed in connection with the detectability of a previously chosen card. Ss of Near Eastern origin tended to show lower pulse rate, higher basic skin conductance, and lower relevant GSR reactivity. It is proposed that psychocultural differences are likely to be the explanation for the consistent differential GSR reactivity to the relevant stimuli.

Attention has recently been drawn to the significant differences in skin conductance found in different ethnic and racial groups. Johnson and Corah (1963) reported that Negro subjects (*Ss*) showed *lower* skin-conductance values than did Caucasian *Ss*. This finding was supported by further data from Bernstein (1965), Johnson and Landon (1965), and Malmo (1965). A study by Lazarus, Tomita, Opton, and Kodama (1966) compared the skin conductance of Japanese and American students as they watched a stressful film. These data suggest that the Japanese may have *higher* skin conductance. More limited data reported by Wenger in Johnson and Landon (1965) as well as by Malmo (1965) suggest that oriental *Ss* may have *lower* skin conductance than Caucasian *Ss*.

Less interest has been focused on analogous differences in GSR reactivity. Sternbach and Tursky (1964) reported skin-potential differences related to ethnic background during the use of electric shock stimulation in connection with pain-threshold determination. Johnson and Landon (1965) reported racial differences in GSR reactivity to the initial presentation of a tone stimulus, but this difference became insignificant upon correction for prestimulus base-line conductance. Lazarus et al. (1966) did not analyze their data on a time scale comparable to the usual GSR analysis, but did report that "Unlike American samples, the Japanese did not show marked stimulus-controlled variation in skin conductance during the stress film . . . [p. 631].

In this laboratory, differences in GSR reactivity were found to exist between Jewish *Ss* of different ethnic backgrounds (Kugelmass, 1963; Kugelmass & Lieblich, 1966). Jewish *Ss* of Near Eastern origin (born in the Moslem countries of the Mediterranean basin, Morocco, Yemen, Iraq, or Iran) showed less GSR reactivity than Jewish *Ss* born in America, Europe, or Israel. These differences in GSR reactivity were observed in connection with the response to the name of the relevant chosen card during the subsequent interrogation part of a card-detection experiment. The purpose of the present report is to present supplementary analyses of the above-mentioned data as well as to report new data derived from independent samples.

METHOD

Original Sample

This sample has been previously described in Kugelmass (1963) and Kugelmass and Lieblich (1966). The Middle Eastern *Ss* exhibited less GSR reactivity in the three different stress conditions, reaching statistical significance in one of them. In addition, 8 of the 36 *Ss* turned out to be nonreactors in all three experimental tests. All eight of these total nonreactors were of Near Eastern origin, and no Western *S* showed this pattern. It was of interest that no significant difference in GSR reactivity was observed in the responses to the other irrelevant cards.

Base-line data. (*a*) Conductance: Values of skin resistance (ohms) were taken 30, 60, 90, and 120 sec. after the beginning of the first 3-min. rest period which opened each experimental examination.

[1] The research reported in this paper has been sponsored by the Air Force Office of Scientific Research, through the European Office, Aerospace Research, United States Air Force, under Contract AF 61 (052)-839.

These values were converted to conductance in micromhos, and averaged for each *S*, and then averaged over *Ss*. The mean conductance value for the Near Eastern *Ss* was 19.74 micromhos, and 17.38 micromhos for the Western group. The difference was not statistically significant at the .05 level ($t = 1.26$ for 13 Western *Ss* and 23 Near Eastern *Ss*). Examination of the data of the subgroup of total nonreactors showed a mean basic conductance of 22.40 micromhos which is significantly higher ($t = 4.10$, with a combined variance estimate of 7.40, $\alpha < .01$).

(*b*) Pulse rate: Five-second samples of pulse-rate data were taken at 30, 60, 90, and 120 sec. after the beginning of the first 3-min. rest period. The mean pulse rate of the Near Eastern *Ss* was 80.49 beats/min, while the rate for the Western *Ss* was 86.45 beats/min. With a combined variance estimate of 145.36, this difference between means is not significant at the .05 level ($t = 1.43$).

It should be added that no significant difference in years of education or intelligence stanine scores was found to exist between the two groups of *Ss* who had a mean number of 6.30 yr. of education ($S = 1.75$).

Blood-Pressure Cuff Sample

This sample was also described in Kugelmass and Lieblich (1966). The 40 *Ss* of this additional sample participated in an experiment in which GSR readings during card detection were made both with and without an accompanying inflation of a blood-pressure measurement cuff. The data to be reported here are all from the *GSR condition* which did not include this additional variance and thus are more comparable to the rest of the data to be presented.

GSR differential reactivity data. The mean reactions to the "hot" relevant chosen card were analyzed as in the original sample, and *Ss* were categorized as reactors if their mean response was at least 1,000 ohms. The relation between ethnic background and this reactivity may be seen in Table 1.

A chi-square analysis of the data indicates that a significantly higher proportion of Near Eastern *Ss* are nonreactors, $\chi^2 = 3.30$ ($\alpha < .05$, one-tailed). Once again no significant difference was found in GSR reactivity to the irrelevant cards.

TABLE 1

RELATION BETWEEN ETHNIC ORIGIN AND GSR REACTIVITY TO THE "HOT" CARD

Ethnic origin	R	NR	Total
Western	10	6	16
Near Eastern	8	16	24
Total	18	22	40

Note.—R = reactor; NR = nonreactor.

Base-line data. (*a*) Conductance: Conductance values were obtained as in the original sample. The mean skin-conductance value of the 24 Near Eastern *Ss* was 17.92 with 18.36 micromhos as a mean for the 16 Westerners. The difference between these two means with a combined variance estimate of 7.50 was not significant at the .05 level ($t = 0.50$).

(*b*) No pulse-rate data had been collected in this experiment.

There was no significant difference in education or intelligence stanine between the Near Eastern and Western parts of the sample which had a mean number of 9.30 yr. of education ($S = 2.06$).

Criminal-Suspect Sample[2]

As a follow-up to the two experiments described above a sample of criminal suspects was examined at the central laboratory of the Israel Police Force. The standard card-test procedure used in the above experiments was incorporated into the regular police interrogation procedure of 62 criminal suspects connected with serious offences. The only other difference in procedure derives from using the police apparatus which was a standard three-channel Stoelting polygraph. This limited the analysis in that no estimate of skin conductance was obtained and changes in millimeter deflections serve as an indication of GSR reactivity.

GSR differential reactivity data. GSR responses were defined as the maximal change in millimeters from base line within 5 sec. of the presentation of the number. Mean responses to each of the cards were calculated for the Near Eastern *Ss* ($N = 39$) and the Western *Ss* ($N = 19$). The records of four of the *Ss* were technically inadequate. The mean response of the Near Eastern *Ss* to the relevant card was 12.51 mm. while for the Western *Ss* it was 21.47. With a combined variance estimate of 303.11 the difference between means is significant at the .05 level ($t = 1.82$, one-tailed). No significant difference was found in GSR reactivity to the irrelevant cards (Near Eastern, $\bar{X} = 7.68$, Western, $\bar{X} = 7.56$, combined $S = 28.91$).

Base-line data. (a) This Stoelting model does not provide for measuring basic conductance.

(*b*) Pulse-rate data: The pulse-rate data were obtained by counting the spikes in the blood-pressure cuff recording which is a standard feature of this apparatus. Five-second samples of pulse-rate data were measured at 30, 60, 90, and 120 sec. after the beginning of the card-test interrogation during the use of the blood-pressure cuff. The mean pulse rate of the Near Eastern *Ss* was 95.99 beats/min, while the rate for the Westerners was 101.90. With a combined variance estimate of 364.07, this difference between the means is not significant at $\alpha = .05$ ($t = 1.13$, 39 Near Easterners, 20 Westerners).

[2] The authors wish to thank the Israel Police Force and especially M. Kaplan, A. Ben-Ishai, and A. Opatowski for their help and cooperation. They also wish to thank G. Ben-Shakhar for expert assistance.

Bedouin Sample

Following the previously mentioned experiment which showed the same general trends, it was decided to obtain a sample of *S*s who approximate the phenotype of the Near Eastern natives as closely as possible. Such *S*s were available in Bedouin Moslem tribes which dwell in the Israeli Negev desert. A detailed description of these Bedouin tribes may be found in Groen, Balogh, Levy, and Yaron (1964). It has been noted that there exists a rough gradient of decreasing modern influence on Bedouin culture in moving from the northern Negev to the southern part where few Jewish settlements exist as yet.[3] As a first step it was decided to draw a sample which would be at least roughly comparable to the previous samples in having some minimum education in a formal setting.

Subjects. Ten male Bedouin *S*s of the Arab El Houzaiel tribe who live close to the Jewish Kibbutz settlement Shuval in the Northern Negev were tested. All of these tribesmen were engaged in agricultural work. The mean level of formal education was 5.8 yr. of school.

Procedure. The *S*s were tested with the Stoelting apparatus obtained from the Police Laboratory as well as with the constant-current (20 μa.) battery-operated GSR apparatus used in the original and blood-pressure cuff experiments described above, which employs Beckman electrodes, on the volar surface of the first and fourth fingers of the left hand. Sanborn paste is used to ensure standard contact. The testing took place in the main room of Sheik Souleiman El Houzaiel on a cool January day (17° C). The *S*s were tested using the two apparatus setups in a rotated order. The standard card test (Kugelmass & Lieblich, 1966) was administered in Arabic through the use of an Israeli police interpreter. The research had been presented to the *S*s as part of a research effort of the Hebrew University to investigate physiological responses of different ethnic groups in Israel.

GSR differential reactivity data. All of the 10 Bedouin *S*s were found to be nonreactors by the criterion previously defined. The records tended to be generally "flat" and even the mean response to the hot card was only 187.50 ohms. Because of this, no significant card detection was possible. As another means of examining GSR reactivity it seemed worthwhile to check the "orienting response" of the *S*s to the first buffer item that is used to open all interrogation sequences. The mean orienting response of the Bedouin *S*s was 1000.00 ohms, that of the Near Eastern of the original sample (most comparable in level of formal education) was 1843.48 ohms, and of the Western *S*s was 1730.00 ohms. Neither the difference between the Bedouin and the Westerners (combined variance estimate 2,188,812, $t = 1.04$) nor the difference between the Bedouin and the Near Easterners (3,974,465, $t = 1.04$) was significant at the .05 level.

Base-line data. (*a*) Conductance: The average basic skin conductance of the 10 Bedouin *S*s was found to be 31.39 micromhos. With a combined variance estimate of 79.55, the difference between this mean and the mean of the Jewish Near Eastern *S*s, 19.74, is significant at the .01 level ($t = 3.50$). The Western mean was even lower, 17.38, and with a combined variance estimate of 109.47 is also significantly lower at the .01 level ($t = 3.11$).

(*b*) Pulse rate: Pulse-rate data were obtained from the Stoelting records as described in the criminal-suspect sample. The mean pulse rate of the 10 Bedouin *S*s was 69.50 beats/min. This is significantly lower than the Near Eastern mean, 80.49 ($\alpha < .05$, $t = 2.42$, with a combined variance estimate of 155.78), and the Western mean, 86.45 ($\alpha < .01$, $t = 3.08$, with a combined variance estimate of 205.87).

DISCUSSION

The ethnic differences that caught the authors' attention in the first two experiments were primarily concerned with a particular kind of GSR reactivity to the relevant card. Following the results described above it seemed important to consider the more sustained base-line state of the organism in addition to the relatively transient reactions of the *S* to specific stimuli.

The data indicate that there may be a tendency for the Near Eastern *S*s to have higher skin conductance. This is clearly seen in the Bedouin sample. Johnson and Corah (1963) considered two aspects of the skin in connection with the racial differences they found. They did not consider the skin color per se as a likely explanation, and Johnson and Landon (1965) were not able to find significant differences in the number of active eccrine sweat glands. While Malmo's (1965) more limited data do suggest a negative correlation between skin conductance and pigment concentration, the present data are not in accord with this suggestion since Near Eastern Jewish and Bedouin *S*s are darker on the average than Western Jews. Another possibility suggested by Johnson and Corah (1963) related lower conductance to a thicker stratum corneum. Superficially, at least, the present data are not in accord with this position. The Bedouin group with by far higher skin conductance is the only group of manual workers who might be expected to

[3] Lt. Col. Sasson Bar-Zvi, former military governor of the Negev region, personal communication, 1966.

have a thick, calloused, stratum corneum. Of course a more precise test of this issue is necessary. Of possible greater relevance to the present data is a finding mentioned by Johnson and Landon (1965): K. M. Taylor found that "Oriental *S*s had greater chloride concentration than Caucasian *S*s [p. 328]." The higher conductance finding in the Bedouin may reflect adaptation to some aspect of the immediate environment, which involves the amount of sweat secreted, its chemical composition, or both.

It is also possible to view these conductance differences as reflecting a difference in basic activation. An examination of the pulse-rate data which is another index of activation suggests caution in such an approach. The pulse-rate data would actually suggest the opposite since the Bedouin *S*s had a significantly lower pulse rate and there is some tendency toward the same in the Near Eastern Jews. It is, of course, difficult to evaluate the comparative psychological states of the *S*s during the examinations, but it does seem plausible to assume much less familiarity and possibly greater stress for the Bedouin *S*s. On the other hand, medical research has focused on the lower risk of cardiovascular disease observed in Yemenite Jews and Israeli Bedouin (Groen et al., 1964). This might reflect a low level of cardiovascular response to stress states. Another way of describing these findings would be to extend the notion of "response stereotypy" used in connection with an individual's psychophysiological responses over situations (Lacey & Lacey, 1958) to ethnic-group response stereotypy. The relatively low variance found in the Bedouin records might even suggest a genetic mechanism for such a notion.

GSR reactivity should be considered in the light of the tendency to find higher base-line skin conductance in the Near Eastern *S*s. All three studies dealing with different Jewish samples indicated significant ethnic differences in reactivity to the "relevant" card as opposed to little or no difference to the irrelevant card. In the case of the Bedouin, all GSR reactivity to the stimuli was relatively low although an appreciable "orienting" reaction was elicited. Previous analysis had suggested some relationship between relevant-card reactivity and base-line skin conductance when considering a special nonreactive subgroup of the Near Easterners as compared with the Westerners (Kugelmass, 1963). Correlations were therefore calculated for the relationship between base-line skin conductance and both relevant- and irrelevant-card GSR reactivity for the original and the blood-pressure cuff samples, and presented in Table 2.

TABLE 2

CORRELATIONS BETWEEN BASIC CONDUCTANCE AND BOTH RELEVANT AND IRRELEVANT GSR REACTIVITY

	Base line (conductance) and cold card (ohms)	Base line and "hot" card
Original sample ($N = 36$)	−0.363*	−0.370*
Blood-pressure cuff ($N = 40$)	−0.155	−0.245

* $\alpha < .05$.

All of these Pearson correlations were negative and low, reaching significance only in the original sample. This direction is what may have been expected from a traditional conception of the Law of Initial Values. The recent findings of Hord, Johnson, and Lubin (1964) as well as Johnson and Landon (1965) obtained significant correlations in the opposite direction, when using conductance changes in micromhos. In the light of their findings it is difficult to conceive of an index

TABLE 3

MEAN CONDUCTANCE CHANGES IN MICROMHOS AND VARIANCE ESTIMATES TO THE RELEVANT CARDS IN THE DIFFERENT SAMPLES

Sample	Middle Eastern sample	Western sample
Original		
$\bar{X}$	.523	.630
S^2	.057	.024
Blood-pressure cuff		
$\bar{X}$	.246	.371
S^2	.054	.050
Bedouins		
$\bar{X}$	.216	—
S^2	.111	—

that completely frees the measure of reactivity from the tonic base-line conductance within the GSR channel. Nevertheless, conductance change in micromhos was calculated for the relevant-card data of the original, blood-pressure, and Bedouin samples and is presented in Table 3. The same pattern of differences in reactivity appears when using this transformation.

Given the above findings it is difficult to conceive of the base-line condition being an important factor in explaining the consistent differential GSR reactivity pattern that emerged in all three Jewish samples. There is some possibility that the influence of the base line comes into operation only after surpassing a minimal level of reactivity which is more likely in the greater response to the relevant card. This should produce a higher negative correlation between relevant-card reactivity and base-line conductance than for the irrelevant-card reactivity. The correlations are too low to supply a clear test of this notion, but this should be considered further. On the whole, the data would suggest that this would only be a minor source for the differential effect, which would appear to be a result of psychological factors.

It is possible to conceive of a range of psychological variables that might bring about such a differential reaction. As a conservative position it seems best to consider all possible factors that produce GSR responses such as shifts of focus of attention, concentration, effort, etc. rather than limit theorizing to the specific lie-detection aspects of the experimental situation. This might be differentially activated by any of the demand characteristics of the card-detection task which might vary in meaning for each individual and probably systematically for groups of different cultural background. A study from this laboratory has recently manipulated this differential GSR reactivity by variations in relevance of stimuli to the person (Lieblich, 1966).

In summary, there appear to be at least two issues to be considered in relating ethnic differences to psychophysiological reactivity which should be considered further. The present data support the notion of consistent differences in basic activation as indexed by skin conductance and possibly pulse rate which do not go in the same theoretically expected direction. Possibly related to these base-line differences are consistent differences in short-term GSR reactivity which appear to be more closely related to psychological factors.

REFERENCES

BERNSTEIN, A. S. Race and examiner as significant influences on basal skin impedance. *Journal of Personality and Social Psychology*, 1965, **1**, 346–349.

GROEN, J. J., BALOGH, M., LEVY, M., & YARON, E. Nutrition of the Bedouins in the Negev desert. *American Journal of Clinical Nutrition*, 1964, **14**, 37–46.

HORD, D. J., JOHNSON, L. C., & LUBIN, A. Differential effect of the law of initial values (LIV) on autonomic variables. *Psychophysiology*, 1964, **1**, 79–87.

JOHNSON, L. C., & CORAH, N. L. Racial differences in skin resistance. *Science*, 1963, **139**, 766–767.

JOHNSON, L. C., & LANDON, M. M. Eccrine sweat gland activity and racial differences in resting skin conductance. *Psychophysiology*, 1965, **1**, 322–329.

KUGELMASS, S. Effects of three levels of realistic stress on differential physiological reactivity. AFEOAR Grant 63–61, 1963. Available from Air Force Office of Scientific Research, Washington, D. C.

KUGELMASS, S., & LIEBLICH, I. Effects of realistic stress and procedural interference in experimental lie detection. *Journal of Applied Psychology*, 1966, **50**, 211–216.

LACEY, J. I., & LACEY, B. Verification and extension of the principle of response stereotypy. *American Journal of Psychology*, 1958, **71**, 50–73.

LAZARUS, R. S., TOMITA, M., OPTON, E., & KODAMA, M. A cross-cultural study of stress-reaction patterns in Japan. *Journal of Personality and Social Psychology*, 1966, **4**, 622–633.

LIEBLICH, I. Identification of variables related to the efficiency of detection of information within the organism, using a physiological index. Unpublished doctoral dissertation, The Hebrew University of Jerusalem, Israel, 1966.

MALMO, R. B. Finger-sweat prints in the differentiation of low and high incentive. *Psychophysiology*, 1965, **1**, 231–240.

STERNBACH, R. A., & TURSKY, B. Ethnic differences among housewifes in psychophysical and skin potential responses to electric shock. *Psychophysiology*, 1964, **1**, 241–246.

(Received May 4, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 2, 163–168

EFFECT OF PERCENTAGE AND SPECIFICITY OF FEEDBACK ON CHOICE BEHAVIOR IN A PROBABILISTIC INFORMATION-PROCESSING TASK [1]

IRWIN L. GOLDSTEIN,[2] JOSEPH T. EMANUEL, AND WILLIAM C. HOWELL

Ohio State University

Proficiency in choosing the better of 2 sources of information was studied as a function of 3 variables: percentage feedback, specificity of feedback, and difficulty of discrimination between sources. 210 Ss performed 200 trials under 1 of 21 conditions. The information pool consisted of 2 distinct distributions of numbers from which random selections were made and displayed as readings on a pair of meters. Only 1 reading was shown per trial, and S judged whether it or the other meter reading (not shown) was higher. Choices could thus be based upon both the momentary state of 1 meter and familiarity with the distributions from which information was drawn. Contrary to typical findings, percentage feedback had no influence on acquisition or performance; paradoxically, a specificity effect appeared when feedback was provided.

The term *knowledge of results* (KR) refers to information provided for an individual or group of individuals concerning the outcome of a response or group of responses. Research on a variety of tasks including the ergograph (Arps, 1917), graded handwriting (Gilliland, 1925), and complex arm-leg coordination (Lindahl, 1945) has demonstrated that performance with KR is superior to performance without KR.

In many realistic settings, however, such as complex military and industrial systems, it is difficult or impossible to provide complete or accurate KR after each response. An issue of some practical importance, therefore, is how often—or on what percentage of responses—feedback must be provided in order to maintain its facilitating influence. Research addressed to this problem has yielded ambiguous results. Kimble (1961), for example, summarizes the findings of several studies concerned directly with percentage of feedback by stating that performance ". . . is usually, but not always poorer . . . [p. 161]" for partial than for 100% conditions. On the other hand, other investigators have reported that performance is not seriously impaired on simple conditioning and reaction-time tasks until reinforcement or feedback is withheld on 70–80% of the trials (McCormack, Bending, & McElheran, 1963; McCormack & McElheran, 1963). Goldstein, Southard, and Schum (1967) obtained a similar result using a complex decision-making task; again, no decrement appeared until KR was withheld on 67% of the decision responses.

Part of the confusion with respect to efficacy of KR may stem from the fact that it is likely to affect more than one organismic process. Bilodeau (1966), for example, distinguishes three potential functions which might be involved—to one or another degree—in any specific instance of KR: a *motivating* function, a *reinforcing* function, and a *directive* function. Each of these processes might be affected somewhat differently by withholding feedback on some proportion of trials. It could be, for example, that the percentage required to maintain a high level of motivation is fixed and relatively small, while that necessary for proper direction varies with such factors as task difficulty, specificity of instructions, and availability of other sources of guidance.

While it is difficult, if not impossible, to isolate one of these KR functions completely in any experimental design, the primary emphasis of the present study was upon the

[1] The research reported in this paper was carried out at the Human Performance Center and was sponsored by the Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, under Contract No. AF 33(615)-2248 with the Ohio State University Research Foundation. Further reproduction is authorized to satisfy the needs of the United States Government.

[2] Now at the University of Maryland.

directive or *informational* facet. Thus, the importance of percentage of KR was examined in light of several other variables which might be expected to control the informational value of feedback provided. These variables were *task difficulty* and *specificity* of KR.

The relevance of task difficulty to the percentage of KR issue is rather obvious. If the information in KR is of any value for response guidance, it should become increasingly important to the subject (*S*) as his course becomes less clear. Consequently, percentage of feedback should be of more significance as difficulty is increased.

On a purely logical basis, specificity of KR also seems to bear directly upon the informational value of feedback; more specific KR should provide more precise guidance of behavior. Actually, however, the relationship is not nearly so simple. Evidence to date suggests that people are limited in their capacity to use highly specific feedback information (Eriksen, 1958; Goldstein et al., 1967); while some specificity may be beneficial, further increases in detail may actually degrade performance (Annett, 1961; Crafts & Gilbert, 1935). Since it is thus difficult to anticipate, for any particular set of conditions, how well specific KR will be utilized, it is virtually impossible to predict the way in which specificity should interact with percentage of KR. Nevertheless, if both factors appear orthogonally in a design, it should be possible to determine whether the specific information is used and, if so, whether it increases in usefulness as it occurs on a greater percentage of trials. This was the approach taken in the present study; two distinct levels of specificity were combined factorially with four percentages of KR and three levels of difficulty.

A task originally developed by Howell and Funaro (1965) provided a useful vehicle for the study of KR variables. In this task, the display consisted of two meters (A and B) upon which could be presented sample numbers (readings) drawn from two distinct number populations. On each trial a sample number was drawn from both A and B, but only one of these was displayed. The *S* was required to guess whether or not it was a higher number than that sampled from the other (not shown) population. These stimulus readings were displayed alternately on the two meters. Thus, as more and more stimulus readings were presented, *S* could gain increasing familiarity with the distributions of numbers from which both the A and B readings were drawn. He could gain even more information about these distributions, however, to the extent that KR was provided and he was able to use it. Difficulty of discrimination could easily be varied by manipulating the differences between the A and B populations.

The Howell and Funaro (1965) study indicated that *S* comes to establish certain numbers as cutoff points for each stimulus distribution such that whenever a reading appears which is below that cutoff point, his choice is the alternate meter. The present investigation examined the effects of percentage of KR (0%, 33%, 67%, 100%) and specificity of KR on *S*'s choice of cutoff points for three levels of input distribution discriminability. Specificity of KR was included at two levels by either permitting *S* to see both the A and B readings after his response to one of them (specific feedback) or by providing him simply with right-wrong information (nonspecific feedback).

METHOD

Subjects. The *S*s were 210 male undergraduate students enrolled in an introductory psychology course. They were divided at random into 21 groups of 10 each. Each *S* served for 200 trials.

Apparatus. The *S* was seated at a table in front of the display and response keys. The room was completely separated from the experimenter (*E*). This display consisted of a pair of microammeters with scale markers reading from 0 to 200 microamps in units of 5. The meter on the left was labeled "A" and the one on the right "B." The *E* was able to adjust and present the readings on either dial by adjusting potentiometers. A pair of switches located beneath each dial was used by *S* in making his choice of dials. The response activated an A or B light in *E*'s booth and the choice was then recorded manually. Another set of switches permitted *E* to present feedback by activating either a green light to signify a correct choice, a red light to signify an incorrect choice, or a combination of the appropriate light and the reading on the alternate meter (i.e., whichever one was *not* used to present the stimulus on that trial).

Independent variables. The independent variables investigated in this study were: (*a*) specificity of feedback, (*b*) percentage of feedback, and (*c*) dif-

ficulty of discrimination, or the relationship of the number distributions underlying the two meters.

Specificity of feedback was manipulated by indicating to *S* either the correctness of his choice alone or the correctness of his choice plus the reading on the alternate meter. Percentage of feedback was varied by presenting feedback either 33%, 67%, or 100% of the time. In order to assess more fully the effects of percentage of feedback, a 0% feedback condition was also included, but as a control condition rather than as part of the overall design.

The difficulty of the discrimination between the two populations was manipulated by varying the means of the two populations. Each population represented a normal distribution of numbers with an *SD* of 5 units. The range of each distribution was 32 units. The mean of the first population (Meter A) was 16 and the means for the second population (Meter B) were either 17, 19, or 21. This separation of 1, 3, or 5 units resulted in the following theoretical differences between the two distributions: For the 1-unit separation the population with the larger mean was higher on 56% of the draws; for the 3-unit separation the population with the larger mean was higher on 67% of the draws; and for the 5-unit separation it was higher on 78% of the draws. The readings (numbers) to be used on any trial were drawn randomly and independently from the two populations; one number from each population (A and B) constituted a trial. One restriction placed on the sampling procedure was that there be no tied readings; whenever this occurred, that pair was discarded and a new pair was drawn. The sampling procedure was carried out by a computer program which also supplied data sheets with all pairs of readings listed for *E*'s use in manual presentation of the readings.

Procedure. The *S* was seated in front of the display and corresponding response keys. He was instructed that the two meters represented two distinct normal distributions and that each trial consisted of a random draw of two numbers, one from each of the two distributions. He was told that one number would be presented to him and the other number would not, and that after each trial both numbers would be returned to the original populations. He was also informed of the type and percentage of feedback under which he would be operating, and the manner in which he was to respond. On each trial, *E* set the value of the potentiometers and presented the reading on one of the meters. After *S* responded, *E* recorded the response and then presented the feedback required by the experimental condition. On successive trials the value presented appeared alternately on Meters A and B so that *S* would have equal opportunities to sample both meters. Each presentation and subsequent recording of the response took approximately 10 sec. Each *S* was required to make 200 choices in a single session which lasted approximately 1 hr. After every 50 trials, summary feedback was provided and a 2-min. rest period was introduced.

RESULTS AND DISCUSSION

For this task, *S* could optimize his prediction by choosing A when the presented value on A was higher than the average value for B or when the presented value on B was lower than the average for A. On all other occasions he should pick B. To determine how closely he approximated this strategy, it was necessary to calculate the empirical cutoff points at which *S* switched his prediction from the presented value on one dial to the unknown value on the opposite dial. The point chosen for analysis in the present study was, for the A-distribution cutoff, the point midway between the highest B value at which A was chosen and the lowest B value at which B was chosen; for the B distribution, it was midway between the highest A value at which B was chosen and the lowest A value at which A was chosen.

In addition to cutoff points, the more usual dependent variable of correct response frequency was also tabulated. Since the latter scores merely confirmed everything shown by the former, they were omitted from the present report.

Average cutoff points obtained under the three levels of input discriminability are shown as deviations from the true distribution means in Figure 1. It is apparent that performance generally improves over blocks of trials; that is, the average cutoff points show progressively less deviation from the true means of the distributions, *F*(7/

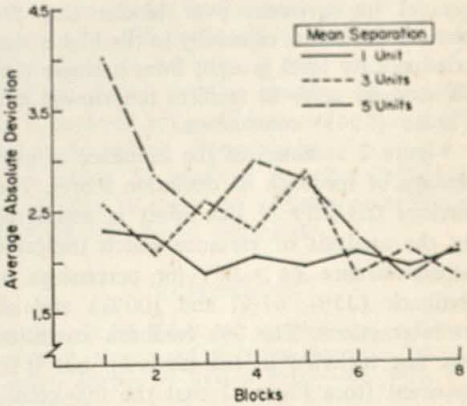


FIG. 1. Average discrepancy between empirically determined and theoretically optimum cutoff points for three levels of input discriminability.

1134) = 14.00, $p < .001$. Also apparent from Figure 1 is the fact that as the two meter distributions become more discriminable, the amount of improvement in performance increases; that is, when the two distributions are widely separated, performance improves much more noticeably over blocks than when the two distributions are closely aligned. This significant interaction of trial blocks with discriminability, $F(14/1134) = 5.85$, $p < .001$, warrants further comment. The performance of the low-discriminability group is initially quite good and shows very little improvement after the early trials, appearing to approach an asymptotic value of 2.0 deviation units during the third block of trials. At this discriminability level the two distributions overlap enough to permit *S* to operate at a near-optimal level by adopting a *single* cutoff for the two distributions. He appears to establish such a cutoff point very early and, having little to gain by adopting a two-cutoff strategy, stabilizes on this value. At the higher levels of discriminability, on the other hand, adoption of a single cutoff value must result in an appreciable reduction in score, and *S* wisely abandons this strategy in favor of the more rewarding double-cutoff approach. Obviously, however, a number of samples must be drawn before he is in a position to establish separate A and B cutoffs; thus, it is not surprising that he initially performs rather poorly but shows rapid and marked improvement as he gains information upon which to base a discrimination. This general improvement over blocks and the obvious deviation, especially in the higher discriminability level groups, from a single cutoff strategy serve to reaffirm the Howell and Funaro (1965) conclusions.

Figure 2 summarizes the influence of percentage of feedback on deviation scores. The obvious triviality of this effect is supported by the analysis of variance which indicated nonsignificance ($p > .05$) for percentage of feedback (33%, 67%, and 100%) and all its interactions. The 0% feedback condition was not included in the analysis, but it is apparent from Figure 2 that the 0% condition does not differ from any of the other conditions. It seems especially surprising that the 0% feedback group was able to approximate an optimal cutoff strategy as well as the 100% feedback group. It is important to note, however, that the 0% feedback group does receive information about each distribution through the alternate presentation of readings on Dial A and Dial B on successive trials; this information, if used, is adequate to permit inference of an optimum strategy.

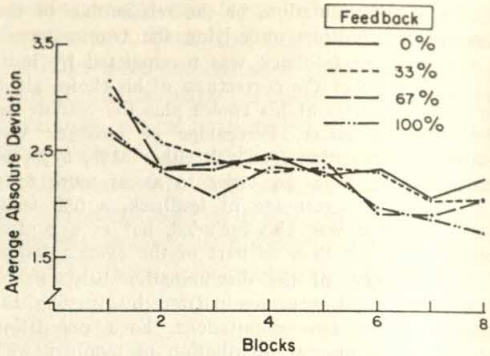


FIG. 2. Average discrepancy between empirically determined and theoretically optimum cutoff points for the four levels of percent feedback.

The conclusion to which one is led by the percentage of KR results, then, is that, contrary to the growing weight of evidence in support of a critical 20–33% KR level, KR can indeed be completely superfluous in maintaining performance. Ammons (1956) noted that in most tasks *S* can find *some* guideline to use in establishing a response criterion. If the stimulus situation actually contains enough information to enable him to perform *optimally*, the response criterion which he sets for himself may be equally as effective as one administered externally. Such, of course, was the nature of the present situation: information provided by the instructions and the alternation procedure seems to have been utilized maximally without the aid of KR.

Since the information available in KR seems generally to have been unnecessary for optimal task performance, it is not surprising that predicted interactions of specificity and difficulty with percentage of KR failed to materialize. This finding strengthens even further the conclusion that KR can be a very impotent directive factor; even when the information is highly specific and task difficulty is very high, it matters little how *much* feedback is provided.

A seemingly paradoxical finding arises, however, in connection with KR specificity. The fact that KR is unnecessary, as shown by the percentage of KR data, suggests that *S* completely *ignores* this source of information. Figure 3 indicates that this is definitely not the case. That he can and does use KR information to improve performance is indicated by the fact that specific feedback conditions were more effective than nonspecific feedback conditions in guiding him toward the optimal cutoffs. Moreover, Figure 3 indicates that the cutoffs established on Meter A were much nearer the optimal cutoff than were the cutoffs established by the same feedback condition groups on Meter B (this would not occur if *S* were ignoring the information completely). The analysis of variance supported both of these apparent differences: both the feedback specificity effect, $F(1/162) = 7.67$, $p < .01$, and the meter effect, $F(1/162) = 11.51$, $p < .001$, were highly significant. The latter finding seems to indicate that *S*s were focusing more of their attention on the left dial than they were on the right dial for reasons that are still unexplained.

It must be concluded, therefore, that although *S* will rely to a degree upon KR information when it is available, he will benefit from it only to the extent that it represents some critical level of specificity. In the absence of KR, he seeks other sources of guidance which, it would seem, may be equally as useful in directing behavior. This conclusion supports the notion of a critical region of specificity for KR (e.g., Ammons, 1956; Annett, 1961), and also demonstrates man's resourcefulness in gathering information upon which to base future responses.

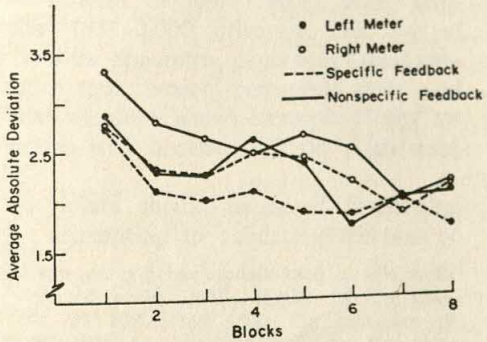


FIG. 3. Average discrepancy between empirically determined and theoretically optimum cutoff points for specific and nonspecific feedback conditions plotted separately for the two meters.

The latter point is of particular significance for real-life situations in which KR is either difficult or expensive to obtain; in such situations it might be well to review the entire informational picture and to consider alternative ways of distributing information-gathering resources. Such an analysis might well suggest a deemphasis of KR relative to other sources of information.

A final point should be made regarding the implications of these data for real-life situations. While the main emphasis of this report is upon KR as *information*, it might be argued that KR is valuable from a *motivational* standpoint as well. If true, one would expect that percentage of KR would have some influence upon level of performance; in the present task, at least, such was not the case. Perhaps if greater value were attached to correct responses, a percentage of KR effect might emerge, serving as evidence for a motivational influence. A preliminary test of this hypothesis, using monetary payoffs as incentive, also failed to produce such an effect.

In summary, it appears that some traditional concepts regarding the role of feedback in establishing and maintaining performance may require revision. In the present investigation, *S*s without extrinsic feedback maintained a performance level similar to *S*s receiving continuous feedback. Exactly what the groups with 0% feedback learned which enabled them to perform as well as a 100% feedback group remains unanswered despite the hypothesis advanced concerning the alternation procedure. Conceivably, the instructions which *S*s received supplied sufficient information to make the addition of extrinsic feedback completely unnecessary.

REFERENCES

AMMONS, R. B. Effects of knowledge of performance: A survey and tentative theoretical formulation. *Journal of General Psychology*, 1956, **54**, 279–297.

ANNETT, J. The role of knowledge of results in learning: A survey. *US Naval Training Device Center Tech. Rep.*, May 1961, No. 342-3.

ARPS, G. F. A preliminary report on work with knowledge versus work without knowledge of results. *Psychological Review,* 1917, **24**, 449–455.

BILODEAU, I. McD. Information feedback. In E. A. Bilodeau (Ed.), *Acquisition of skills.* New York: Academic Press, 1966. Pp. 255–296.

CRAFTS, L. W., & GILBERT, R. W. The effects of knowledge of results on maze learning and retention. *Journal of Educational Psychology,* 1935, **26**, 177–187.

ERIKSEN, C. W. Effects of practice with or without correction on discrimination learning. *American Journal of Psychology,* 1958, **71**, 350–358.

GILLILAND, A. R. The effect of practice with and without knowledge of results in grading handwriting. *Journal of Educational Psychology,* 1925, **16**, 532–536.

GOLDSTEIN, I. L., SOUTHARD, J. F., & SCHUM, D. A. Feedback in a complex multiman-machine system. *Journal of Applied Psychology,* 1967, **51**, 346–351.

HOWELL, W. C., & FUNARO, J. F. Prediction on the basis of conditional probabilities. *Journal of Experimental Psychology,* 1965, **69**, 92–99.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

LINDAHL, L. G. Movement analysis as an industrial training method. *Journal of Applied Psychology,* 1945, **29**, 420–436.

MCCORMACK, P. D., BENDING, F. R. S., & MCELHERAN, W. G. Effects on reaction time of partial knowledge of results of performance. *Perceptual and Motor Skills,* 1963, **17**, 279–281.

MCCORMACK, P. D., & MCELHERAN, W. G. Follow-up of effects on reaction time with partial knowledge of results. *Perceptual and Motor Skills,* 1963, **17**, 565–566.

(Received May 17, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 2, 169–176

RELATIONSHIP BETWEEN PERCEPTUAL STYLE AND DRIVER REACTION TO AN EMERGENCY SITUATION [1]

GERALD V. BARRETT AND CARL L. THORNTON [2]

Goodyear Aerospace Corporation, Akron, Ohio

Previous investigators have found low relationships between human characteristics and accident behavior. The possibility was raised that these findings might be due to a number of factors including indiscriminate grouping of accidents and choice of predictors which have no logical relationship to accident behavior. In the present study 20 drivers were subjected to a controlled emergency situation in an unprogrammed automobile simulator, where a pedestrian (dummy) emerged from a shed into the path of the vehicle. Subsequent measurements of individual perceptual style, using the Rod and Frame Test (RFT), were correlated with effectiveness of reaction to the emergency situation. Series 3 of the RFT correlated .67 with reaction time, .75 with deceleration rate, and .50 with a hit-miss criterion. Individual differences in perceptual style and simulator discomfort resulted in shrunken multiple correlations of .89–.95 depending on the data transformations. The extrapolation of the results to real-world accident behavior has implications especially with regard to age, sex, and use of alcohol by drivers.

In general, the obtained relationships between various measures of driver characteristics and accidents have been quite low. Goldstein (1961), in an extensive review of research on human variables assumed to be important in safe motor-vehicle operations, found low relationships between various predictors and accident behavior. He noted, for example, in over 45 studies that the correlations between criterion measures and visual functioning (visual acuity, depth perception, eye-muscle balance, etc.) were never above .20. Similarly, correlations were no higher than .17 between reaction time (RT) and accident criteria in four studies using large samples (331–3,000 drivers). The use of psychomotor apparatus, paper-and-pencil psychomotor tests, sensory-perceptual tests, and cognitive measures shows correspondingly low correlation with accident criteria (Goldstein, 1962).

Two points should be noted concerning studies attempting to validate predictors of accident criteria. First, the investigators have used very heterogeneous measures of driver accident behavior. There appears to have been little attempt to categorize or logically group accidents. A tacit assumption seems to be that all accidents have a common cause and therefore can be grouped together. In opposition to this point of view, one could reason that accident behavior is quite diverse and complex, having many causes, and only by an appropriate categorization of accident behavior would it be possible to obtain a better understanding of the phenomenon. Second, the predictors chosen often fail to have any theoretical or logical relationship to the accident behavior which is to be predicted. The failure to categorize accidents logically and/or to choose a predictor which has a theoretical relationship to the category of accidents may account, to a large extent, for the low correlations which have been reported.

An attempt to eliminate the category problem was made by Barrett and Nelson (1965b, 1966b). They tested driver reaction to what could be considered one category of accidents, a situation where the driver was forced to make an emergency stop because a pedestrian emerged suddenly into the vehicle's path. This emergency situation was staged while

[1] The emergency-behavior data used in this study were collected under Public Health Service Contract No. PH-108-64-168 while the collection of perceptual style data was supported by Goodyear Aerospace Corporation.

[2] The authors express appreciation to Margaret J. Smith for statistical advice and to Patrick A. Cabe for computer analysis of the data.

the driver was operating an unprogrammed automobile simulator.

The present study attempted to find a predictor which would have a logical relationship to the emergency responses. An analysis of the situation indicated that the main tasks were to detect and identify the pedestrian. This appeared to be related to Witkin's concept of perceptual style (Witkin, Dyk, Faterson, Goodenough, & Karp, 1962). Witkin classified individuals on a field-dependent–field-independent continuum, along which individuals may be differentiated with respect to their ability to pull a figure from an embedding context, with field-independent people being better able to perform the task. Sudden pedestrian emergence into the field of view presented a figure in an embedded context.

Since the behavior required of the subject (*S*) in the simulated driving situation appeared to be related to perceptual style from this logical viewpoint, it was hypothesized that the field-independent individuals should be more effective in reacting to the emergency driving situation than would field-dependent individuals.

METHOD

Automobile Simulator

An engineering test model of an unprogrammed automobile simulator was the basic research tool used to obtain drivers' reactions to an emergency. Figure 1 shows the arrangement of the simulator in the test area with the horizon panels around the terrain model and the black-out partition between the terrain model and the screen removed. The terrain model was an 87:1 (HO gauge) scale representation of several flat roads including a wide road beside which a small shed was located. This shed contained a pedestrian dummy, scaled to 5.5 ft., which could be released onto the roadway when a

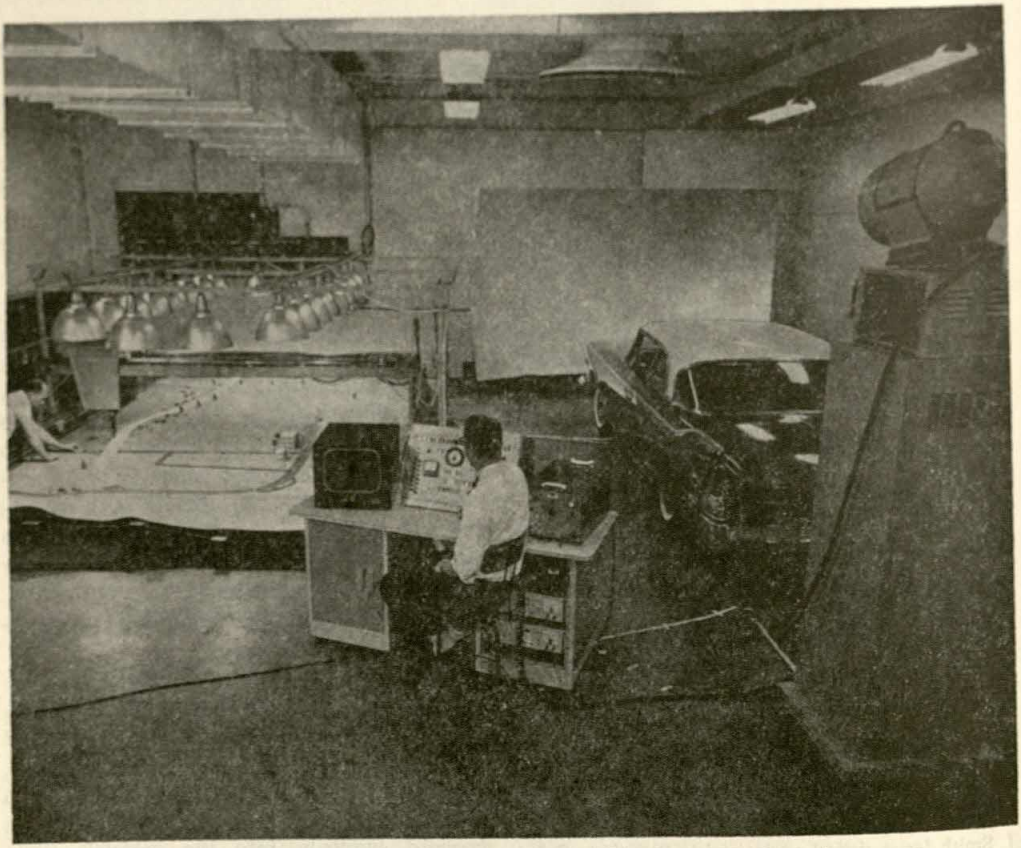


FIG. 1. Automobile simulator, in projected image display mode, with horizon panels and black-out partition removed.

microswitch was tripped by the television pickup carriage. A television camera mounted on a transport system viewed the terrain model from above through a special lens assembly and moved in direct response to the movements of the automobile's control elements.

A projected-image visual display was used with half the *Ss* and a virtual image display with the remainder. Both visual displays gave the driver approximately a 50° horizontal angle of view. For the purposes of this study the two displays were considered to be functionally equivalent. With each visual display the driver sat in the automobile and performed the usual control movements associated with driving. A more detailed description of the simulator and associated visual displays has been reported by Barrett and Nelson (1965c, 1966a).

During the emergency trial a continuous measurement of speed, time, brake position, steering-wheel position, lateral position of automobile, longitudinal position of automobile, and pedestrian position was recorded. From these measures, effectiveness of driver response to the emergency was obtained.

Rod and Frame Apparatus

Perceptual style was measured with the standard Rod and Frame Test (RFT) described by Witkin, Lewis, Hertzman, Machover, Meissner, and Wapner (1954). The apparatus consisted of a luminous square frame (40 in.), pivoted at its center so that it could be tilted left or right. Pivoted at the same center, but moving independently of the frame, was a luminous rod (39 in.).

The *S* was seated in a chair 8 ft. from the rod and frame that could be placed in three positions: erect, tilted 28° to the left, or tilted 28° to the right. The above apparatus was situated in a completely darkened room and *S* wore goggles with dark lenses so that only the rod and frame could be seen.

Subjects

A total of 50 male *Ss* between the ages of 30 and 45 were selected at random from approximately 1,200 employees in an engineering division of an aerospace corporation. Twenty-six *Ss* left the simulator because they experienced extreme discomfort (simulator sickness). The data for two *Ss* were not used because of an error in procedure, and the data for one *S* were not used because he became aware of the purpose of the study. The remaining 21 *Ss* completed the emergency-behavior trials. One *S* was not available for subsequent testing with the RFT so that the results in this study were based upon the driving behavior and perceptual style responses of 20 *Ss*.

Procedure for Emergency-Behavior Study

The same procedure was followed for all 20 *Ss* consisting of three orientation drives around the terrain model and a pretrial run for a study of driving at requested speed without knowledge of apparent speed. The complete procedure for the speed study has been described by Barrett and Nelson (1965a).

Upon making the tenth and final speed judgment, *Ss* followed the same procedure as in the previous nine trials. They were not told the number of speed trials to be made nor that the last speed trial had been completed. The *Ss* made left U turns at the end of the roadway and proceeded down the middle of the right 8-ft. lane in the emergency-behavior test trial. As in the preceding trials, the device operator monitored speed at the control panel and either instructed the driver to maintain, increase, or decrease his speed so that a speed of approximately 25 mph was maintained. The speedometer in the vehicle had been covered by tape at the beginning of the speed trials and remained covered for the emergency trial so that the instrument panel was not a distraction.

While the driver was approaching the dummy, the experimenter (*E*) stopped giving speed instructions. When the driver reached the release position (76.5 ft. equivalent from front bumper to the shed) the pedestrian emerged at a controlled rate from the shed and the driver's responses were recorded.

Procedure for Measuring Perceptual Style

Approximately 6 mo. after the data had been collected for the emergency-behavior study, the perceptual style of *Ss* was measured with the Rod and Frame Test (RFT). Measurement followed the standard procedure developed by Witkin et al. (1954) consisting of three series of eight trials each. Series 1 consisted of trials in which *S* was tilted 28° in the same direction as the tilt of the frame; Series 2 was trials in which *S* was tilted 28° in the direction opposite that of the tilt of the frame; Series 3 consisted of four trials with the frame tilted 28° to the right and four trials with the same tilt to the left while *S* remained erect. The *S*'s task was to position the rod to what he considered to be true vertical by asking *E* to move the top of the rod right or left. The *S*'s score was the mean degrees error on each series. In addition to mean scores on each of the three series, an Orientation Index was computed from the standard scores of the three series as outlined by Witkin et al. (1954).

Procedure for Measuring Discomfort in the Simulator

Since a certain degree of discomfort was experienced even by those who completed their trials in the driving simulator, two measures of discomfort were obtained after *Ss* completed the RFT. The first was a simple graphic rating scale on which each *S* rated his perceived discomfort, from no discomfort (0) to extreme discomfort (10). The second was *S*'s estimate of how long after he left the simulator he still felt discomfort.

TABLE 1

INTERCORRELATIONS OF EMERGENCY BEHAVIOR WITH RFT MEASURES

	Reaction time	Log reaction time	Deceleration rate	Deceleration rate squared	Log deceleration rate	Hit-miss
Series 1	.24	.21	−.13	−.09	−.17	−.01
Series 2	.37	.35	−.36	−.32	−.41	.16
Series 3	.61**	.57**	−.55**	−.53*	−.57**	.34
Orientation index	.34	.39	−.39	−.35	−.42	.19
Log Series 3	.67**	.63**	−.69***	−.69***	−.70***	.45*
Reciprocal Series 3	−.63**	−.60**	.74***	.75***	.74***	−.50*

* $p < .05$.
** $p < .01$.
*** $p < .001$.

RESULTS

The results of the perception test (RFT) were compared to (1) initial brake reaction time, that is, the time which elapsed from the emergence of the dummy until *S* began to depress the brake pedal; (2) deceleration rate, that is, the miles per hour per second which the apparent motion of the simulator was decreased by *S* as measured from the time of emergence of the dummy to either subsequent impact with the dummy, or the obtaining of zero velocity before collision; and (3) the hit-miss dichotomy found in most real-world studies. Series 1, 2, 3, and the Orientation Index from the RFT were compared with these three criteria.

The results confirmed the hypothesis that perceptual style was significantly related to the ability to react to emergency situations. However, only one aspect of the RFT correlated with the criteria.

Table 1 shows that correlations with Series 1, 2, and Orientation Index were not significantly related to emergency behavior. Differences between these aspects of the test and Series 3 will be discussed later in connection with Witkin et al.'s (1954) findings.

The correlation (Pearson r) between Series 3 of the perception test and reaction time in the simulator was .61 ($p < .01$). The deceleration rate ($r = -.55$, $p < .01$) was also significantly related to Series 3. However, the relationship between the hit-miss dichotomy and Series 3 was only .34 (nonsignificant) using a point-biserial r. This latter finding emphasizes the importance of the interval measurement used in this study.

Inspection of the data revealed that the relationships were not linear. It became evident that the empirical data were more nearly approximated by a log function for reaction time (see Figure 2) and a hyperbolic function for deceleration rate (see Figure 3) and the hit-miss dichotomy. Correlations increased in magnitude from .61 to .67 for reaction time, from .55 to .75 ($p < .001$) for deceleration rate, and from .34 to .50 ($p < .05$) for the hit-miss dichotomy. A single perceptual measure, then, was able to account for 25–56% of the individual differences in ability to react to the emergency situation.

It was apparent from the data analysis that discomfort, due to the experience of operating the simulator, affected the results. Multiple regressions were calculated using a Wherry Test Selection Program with F test (Wherry & Smith, 1966) for the IBM 1620

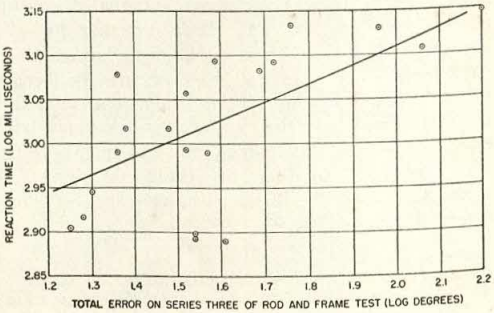


FIG. 2. Relationship between perceptual style and reaction time.

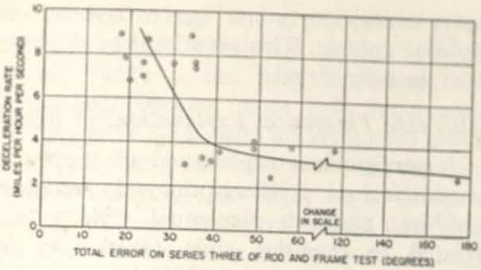


FIG. 3. Relationship between perceptual style and deceleration rate.

computer. It was found that discomfort was not a significant factor in initial reaction time, but was quite important for deceleration rate. Multiple Rs obtained ranged from .89 to .95 ($p < .001$) depending on the particular data transformations involved. Table 2 shows the multiple Rs obtained using values of x, x^2, $\frac{1}{x}$, and log x as units representing various curvilinear functions that could yield a data fit. Two factors, perceptual differences and simulator discomfort, accounted for 79–90% of the variance in deceleration rate of the simulator.

As mentioned previously, high correlations were found between emergency behavior and Series 3 only. Previous research (Witkin et al., 1954) showed that the RFT was not a unitary test. Series 1 and 2 were found to be similar, while Series 3 (no body tilt) tapped a different dimension of perception. Embedded Figures Test (pulling an item from an embedded context), often used as a measure of perceptual style, was found to be most highly related to Series 3 and not to Series 1 and 2. Witkin's perception tests were found to be on a difficulty continuum. The Embedded Figures Test and Series 3 were the easiest of the perceptually embedding tests. Since at the time of the experiment it was not known how the driving task would compare to the difficulty levels of perceptual style, all three series of the RFT were given.

TABLE 2

VARIABLES APPEARING IN THE MULTIPLE-REGRESSION EQUATION USING VARIOUS DATA TRANSFORMATIONS—DECELERATION RATE AS RELATED TO PERCEPTUAL STYLE AND DISCOMFORT

Item	Variable					Multiple R
	1	2	3	4	5	
Deceleration rate (DR)	$\frac{1}{S_3}$	D^2	$\log S_1$			.90*
$(DR)^2$	$\frac{1}{S_3}$	$\log S_1$	$\frac{1}{I_A}$			.89*
$\frac{1}{(DR)}$	$\frac{1}{S_3}$	D^2	$\frac{1}{S_1}$	$\frac{1}{D}$	S_2^2	.95*
$\log (DR)$	$\frac{1}{S_3}$	D^2	$\log S_1$			.90*

Note.—S_1 = Series 1 of the Rod and Frame Test (RFT); S_2 = Series 2 of the RFT; S_3 = Series 3 of the RFT; D = rating of subjective discomfort during simulator operation; I_A = estimate of the length of time the discomfort lasted after leaving the simulator.
* $p < .001$.

DISCUSSION

The present results support the hypothesis that field-independent individuals are more effective in responding to emergency situations. The finding of high relationships, accounting for approximately 80% of the variance, was probably a function of a number of interrelated factors.

Five such factors were considered to be important in obtaining significant results in this type of accident research: the studying of only one specific type of accident, using an improved research tool to study a specific behavioral sequence, unitary exposure to the emergency sequence, the continuous measurement of relevant criteria, and the utilization of a predictor with a logical relationship to the criterion.

The importance of studying one type of accident is shown by noting that such accidents as rear-end collisions, loss of control on turns, accidents during rain or snow, and accidents related to mechanical malfunction are usually grouped together. They may have completely unrelated human characteristics associated with incidence and severity of outcome. Also, the classification of real-world accidents often does not take into account the problem of human culpability. The tacit assumption is that an accident is the driver's fault. The term "accident" is often a legal designation rather than a description of a behavioral act. It is quite possible that

having an accident by the legal definition is the most effective driving behavior (e.g., hitting a parked car to avoid a child).

Significant findings have often been the result of improved research tools. Looking at a specific structure rather than just a global process has made it possible to confirm or refute many theoretical issues. By using an automobile simulator, control and continuous measurement of pertinent variables in an emergency situation were possible. Also, the simulated situation presented every driver with the same emergency, giving constant exposure across *Ss*. This is not to imply that the simulator used in the present investigation is not without flaws, especially as evidenced by the discomfort experienced by the majority of *Ss*.

The predictor in this study had a logical relationship to the criterion behavior. Often simple brake reaction to a red light is correlated with accident records (e.g., Cation, Mount, & Brenner, 1951) where there is little correspondence between the braking response elicited by the red light and the complex perceptual process in the real world.

Related Real-World Research

In a review of the literature, one real-world study was found which appeared to be related to the behavioral sequence analyzed in this experiment. An attitude scale, picture-frustration test, Raven Progressive Matrices, and a visual perception test were administered to accident repeaters and accident-free drivers (Spicer, 1964). Only the visual perception test, motion picture films representing a wide variety of traffic conditions, was consistently related to driving and traffic mishaps. The *Ss* checked statements concerning each of 11 film sequences to indicate what he felt was important to the driver of the car. This test might have some relationship to the concept of perceptual style since *S* was required to extract the salient information from the film strip in a short period of time (35 sec.).

It was interesting to note that adolescents did not perform as well as older drivers on the visual perception test. This could have been a function of either inexperience and therefore lack of awareness of vital cues in driving or field dependence and consequent lower ability to pull a figure from an embedding context. This point will be discussed later in more detail.

Alternate Theoretical Explanations of Results

A parsimonious explanation of the results would include a description of perceptual abilities, although perceptual style is also related to cognitive style (Witkin et al., 1962). Field-independent individuals tend to think more analytically and critically. Possibly, field-independent individuals reacted faster because they already had a preformed set to respond, thinking out in advance how they would react to an emergency situation. In contrast, the field-dependent person, being less analytical, may never have given any consideration to how he would react.

Witkin et al. (1962) have reported that field-dependent individuals become confused, disorganized, and inadequate under stress. Since confusion could also be a factor in emergency situations, there apparently are a number of possible alternative explanations using other correlates of perceptual style which could account for the less effective behavior of the field-dependent *Ss*.

Extrapolation of Results to Real-World Accident Situations

Three complex real-world driving behaviors can be explained on the basis of the results of this investigation and other related perceptual style research. First, it has been well established that alcohol is a factor in many accidents (Fox & Fox, 1963). Witkin et al. (1962) report that perceptual style, as measured specifically by *S*'s ability to visually extract a figure from an embedding context, is affected by alcohol. It is conceivable therefore that while under the influence of alcohol a driver would find it harder to extract an emerging pedestrian, vehicle, or other obstacle from the complex visual field. This could be one factor in explaining why drivers under the influence of alcohol are involved in a high percentage of accidents.

Second, studies are consistent (Goldstein, 1962) in showing that young drivers (below 25) and older drivers (beyond 65) have a disproportionately high rate of accidents. It has also been shown that perceptual style is

a function of age. In a number of cross-sectional and longitudinal studies (Witkin et al., 1962) it has been determined that individuals become more field independent as they grow older, with an asymptote around 21. Since young drivers (16–21) are relatively field dependent, this might be a factor in their higher accident involvement. While the younger drivers as a group have a faster simple reaction time, the real-world situation involves a complex perceptual element and the intruding object must first be recognized before an individual can respond to it. Similarly, recent evidence (Schwartz & Karp, 1967) indicates that older individuals are field dependent.

Third, previous research (Witkin et al., 1962) has shown that women as a group are more field dependent than men. On that basis one would predict that women drivers who are presented with an emergency driving situation would be less effective in responding than men. One real-world study has a possible bearing on this point. Uhr (1959) found that women were more likely to cut across a motor scooter's path from a stop street than were men. It is possible that the women were unable to perceptually extract the small scooter from the background as quickly as the men.

It appears feasible to extrapolate the results of the present study to several other real-world situations. Any occasion where an individual must quickly detect, identify, and properly react to a stimulus could be related to perceptual style. Many examples could be given such as sudden emergence of a child from between parked cars, the swerving of an automobile into the driver's projected path, or the emergence of an automobile from an alley or hidden driveway. The generality of these results, however, must await future research.

Past research with the Embedded Figures Test (EFT) and the RFT may also be brought to bear on practical traffic problems. It is known that various factors decrease one's ability to pull an embedded item from its context, such as alcohol, fatigue, and brain damage. Besides these factors, which may cause severe decrement in performance for anyone, there are subtle genetic and environmental factors which cause a wide range of individual differences even under optimal testing (or driving) conditions. For this reason the EFT and RFT when properly validated could be used to screen large populations of drivers who either are planning to enter or are presently employed in an occupation which requires a great deal of driving (such as truck, bus, and taxi drivers, or traveling salesmen).

Since it is not now feasible to restrict field-dependent people from driving, training could be established for those who find driving to be an occupational necessity, but are extremely field dependent. Perception, per se, is resistant to training. It is, however, possible to "train around" certain deficits. That is, it is possible to identify perceptual deficits and teach people to compensate for them. One such possibility is to train field-dependent persons to respond completely (i.e., reach full brake depression), as soon as they detect the emerging image. In this manner, the lag time between detection and identification may be utilized in decelerating the vehicle. This amounts to telling the driver to "be careful," which many slogans do, but in a very specific way.

In conclusion, this discussion illustrates one of the main advantages of physical stimulation of the driving task. Besides factors such as exact experimental control and safety, confirmation of hypotheses which could not be easily evaluated in the real world is possible. Further research, of course, is required to confirm the hypothesized functional relationships between perceptual style and human performance in specified emergency situations. Also, a replication of this study in the real world should be conducted, especially in light of the restriction in sample imposed by the "simulator sickness." There is now, however, some empirical evidence which gives a basis for a number of specific investigations.

REFERENCES

BARRETT, G. V., & NELSON, D. D. Driving at requested speed. Volume 12, Goodyear Engineering Report-12400, Contract No. PH-108-64-168, December 1965. (a)

BARRETT, G. V., & NELSON, D. D. Emergency behavior. Volume 13, Goodyear Engineering Report-

12400, Contract No. PH-108-64-168, December 1965. (b)

BARRETT, G. V., & NELSON, D. D. Introduction to driving simulator evaluation. Volume 1, Goodyear Engineering Report-12400, Contract No. PH-108-64-168, December 1965. (c)

BARRETT, G. V., & NELSON, D. D. Description of virtual image display. Volume 16, Goodyear Engineering Report-12400, Contract No. PH-108-64-168, March 1966. (a)

BARRETT, G. V., & NELSON, D. D. Emergency behavior study with virtual image display. Volume 19, Goodyear Engineering Report-12400, Contract No. PH-108-64-168, March 1966. (b)

CATION, W. L., MOUNT, G. E., & BRENNER, R. Variability of reaction time and susceptibility to automobile accidents. *Journal of Applied Psychology,* 1951, **35**, 101–107.

FOX, B. H., & FOX, J. H. (Eds.) *Alcohol and traffic safety.* Washington, D. C.: Public Health Service Publication No. 1043, 1963.

GOLDSTEIN, L. G. Research on human variables in safe motor vehicle operation: A correlational summary of predictor variables and criterion measures. The Driver Behavior Research Project, George Washington University, 1961.

GOLDSTEIN, L. G. Human variables in traffic accidents: A digest of research and selected bibliography. National Academy of Sciences–National Research Council Publication 1054, 1962.

SCHWARTZ, D. W., & KARP, S. A. Field dependence in a geriatric population. *Perceptual and Motor Skills,* 1967, **24**, 495–504.

SPICER, R. A. Human factors in traffic accidents. United States Public Health Service Research Grant No. AC-55, Final Report, May 1964.

UHR, L. Sex as a determinant of driving skills: Women drivers. *Journal of Applied Psychology,* 1959, **43**, 35.

WHERRY, R. J., & SMITH, M. J. *Wherry test selection program with F-test.* Columbus: Ohio State University, 1966.

WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., & KARP, S. A. *Psychological differentiation: Studies of development.* New York: Wiley, 1962.

WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

(Received June 9, 1967)

journal of applied psychology

VOL. 52, NO. 3 JUNE 1968

EFFECT OF LUNAR GRAVITY ON MAN'S PERFORMANCE OF BASIC MAINTENANCE TASKS

RICHARD J. SHAVELSON AND JOSEPH L. SEMINARA [1]

Lockheed Missiles and Space Company, Sunnyvale, California

9 Ss were trained extensively on 3 maintenance tasks: bolt torquing, connector mating, and nut threading. They were randomly distributed into 1 of 3 clothing conditions (shirt-sleeve, vented suit, and pressurized suit) and trained and tested on all 3 tasks in 3 gravity conditions (1 gravity, 1 gravity harnessed, and 1/6 gravity). This study demonstrated that lunar gravity imposed a 25% performance decrement over performance in 1 gravity ($p < .01$). The vented suit imposed a 60% performance decrement ($p < .01$), and the pressurized suit imposed a 150% performance decrement ($p < .01$) when compared to performance in the shirt-sleeve mode. On the basis of these findings and subjective reports, preliminary human-factors design criteria were suggested for lunar-gravity performance aids. The need for subsequent research in the areas of mission-specific maintenance tasks and candidate job aids to improve performance in the lunar environment was pointed out.

Past research, employing a variety of simulation methods, demonstrated that man's behavior in lunar gravity differed considerably from his behavior in the earth's gravity. Hewes [2] and Spady (1964a, 1964b) found that man could jump vertically as high as 14 ft. in lunar gravity and perform other ordinary tasks in an extraordinary manner. Roberts (1963) investigated man's walking behavior in lunar gravity which he described as a fast walk in slow motion. Prescott and Wortz [3] (1966) and Hazard (1965) showed that, during walking, man's metabolic rate decreased in 1/6 gravity when compared to his 1-gravity rate. Although there has been a comprehensive examination of man's physical and physiological behaviors in lunar gravity, a few studies have concerned themselves with man's performance of maintenance tasks. Holmes (1965) studied the effectiveness of a variety of tools on maintenance tasks in various gravity conditions including lunar gravity. Using two subjects (*Ss*) for his lunar-gravity simulation, he found that performance times increased by 8% over performance in the earth's gravity. However, due to the paucity of *Ss*, the large number of tasks under examination, and the lack of statistical treatment, Holmes' findings were considered tentative, at best.

The purpose of the present research was to develop data regarding the effects of lunar gravity and candidate lunar clothing conditions (shirt-sleeves, vented suit, pressurized suit) on man's performance of three basic maintenance tasks (bolt torquing, connector mating, and nut threading) to serve as guidelines for task analyses, timelines, and performance-aid designs. It was hypothesized that performance times in 1/6 gravity would increase over those in 1 gravity, that performance times would increase with increasingly encumbering clothing requirements, and that the connector-mating task would be the

[1] The authors wish to acknowledge the assistance of G. F. O'Keefe, C. J. Baker, and A. L. Hall in providing life support, communications, and technical advice for pressure suit research, and R. E. Mahan, simulation engineer.

[2] D. E. Hewes, A. A. Spady, Jr., & R. L. Harris. Comparative Measurements of Man's Walking and Running Gaits in Earth and Simulated Lunar Gravity. NASA Tech. Note D-3363. Unavailable.

[3] E. C. Wortz & E. J. Prescott. The Effects of Subgravity Traction Simulation of the Energy Costs of Walking. Personal communication, 1965. Pp. 2–26.

© 1968 by the American Psychological Association, Inc.

least affected by the gravity or clothing-condition variables.

METHOD

Subjects

Nine male *Ss* participated in this study. The *Ss* were selected from a pool of "astronaut" test *Ss* maintained by Lockheed Missiles and Space Company's (LMSC) Biotechnology Organization. The *Ss* fell between the ages of 23 and 36, all were engineers or physical scientists, all were under 6 ft. in height, and all had performed as a pressure-suited test *S* in previous simulation research.

Apparatus

The effect of lunar gravity was achieved by using LMSC's ⅙-gravity simulator, LUNARG. This unit, which supports five-sixths of the *S*'s weight, consisted of nine negator spring motors attached to a cable reeling system. The *S* was attached to the aforementioned cable via a harness. Fine balancing of plus or minus 1 lb. accuracy was achieved by placing a skin diver's weight belt around *S*'s waist.

The space suit employed in this study was an Apollo A-4H training suit, on loan from NASA, Houston. The suit consisted of a Model C3 helmet, a torso-limb assembly, gloves, and astro-cap.

Since the Apollo A-4H pressure suit had a helmet without a visor, the vented-suit condition was defined as helmet and gloves on with an air ventilization flow of 10 cu. ft. per minute. The pressurized-suit condition was defined as suit pressurization at 3.7 ±2 psi differential with air. The clothing condition referred to as shirt-sleeve was defined as street clothing: shoes, socks, slacks, and shirt.

The performance-fixture apparatus was a unistrut frame structure 9½ ft. in height and 19 in. in width on which a uniform-sized panel, 54 in. in height × 19 in. in width × ¼ in. in depth, was placed. The panel contained three tasks: (1) bolt torquing, (2) connector mating, and (3) nut threading. On *S*'s left-hand side of the panel were nine threaded receptacles into which bolts 1 in. long with a 7/16 of an inch hexagonal head were screwed and then torqued down to 70 in. lb. using a torque wrench. Nine identical Deutsch push-pull type connectors (½ in. in diameter) were located in the middle of the panel. Twelve equally spaced bolts of three different sizes (¼, 6/16, ½ in.) protruded from the right-hand side of the panel. Two of the nuts were hexagonal and fitted on the two larger bolts. The third was a small wing nut. Attached to the performance fixture at a height of 43 in., and extending out to the right of the panel, was a contents tray (24 × 6½ × 6½ in.). During the experiment, this tray contained the torquing wrench and a predetermined number of each of the task elements.

Design

The experimental design employed in this research was a partial replication of an 81-celled 3 × 3 × 3 factorial. Nine *Ss* were randomly assigned to one of three clothing conditions (shirt-sleeve, vented suit, or pressurized suit) with three *Ss* in each condition. All nine *Ss*, regardless of clothing condition, performed all three subtasks (torquing, connector mating, and threading) in all three gravity conditions (1 gravity, 1 gravity harnessed, and ⅙ gravity). In order to minimize the effects of practice, *Ss* were tested in the principal condition, gravity, in a counterbalanced order.

All *Ss* received task orientation and preliminary training in the shirt-sleeve, 1-gravity environment. They were then trained and tested in their respective clothing conditions.

Procedures

The experiment was divided into two parts: (1) orientation and preliminary training, and (2) specific training and testing. The former part describes the instructions given to the *Ss* and the base-line training in the shirt-sleeve 1-gravity environment. The latter part describes training in specific clothing and gravity conditions and testing.

Orientation and preliminary training. The *Ss* were instructed to perform each task separately, always working from panel top to bottom. They were told to work as quickly and as accurately as possible since both time and errors were being recorded. One exception to this procedure was necessary. During the course of the experiment, it was discovered that *Ss* participating in the pressurized-suit condition were unable to successfully complete all of the task elements on the performance panel. This was not due to the inability of *Ss* to reach all of the task elements, since for all *Ss* regardless of clothing condition the top of the panel was placed 1 in. above their shirt-sleeve height, but rather to the amount of fatigue and the difficulty of manipulation in the extreme high and low positions. For example, although *Ss* were able to place the nine bolts into the panel, they were unable to torque the bolts in the highest and lowest positions with the torque wrench to the preset value of 70 in. lb. As a consequence, the work area of the panel was reduced such that the pressure-suited *Ss* were required to complete the middle five torquing and connector-mating task elements, and the middle eight threading task elements. Average time to complete one task element was computed for all *Ss* on all tasks.

At the completion of the instructions, *Ss* were given an opportunity to manipulate the three tasks until they were satisfied that they had fully understood the instructions. The *Ss* were then trained in the shirt-sleeve 1-gravity mode until an asymptote in time and errors was reached. This required about one-half to 1 hr.'s time.

Specific training and testing. Specific training and testing proceeded first with the shirt-sleeve group, then with the vented-suit group, and lastly with the pressurized-suit group. In all three clothing conditions, *Ss* were trained to asymptote in time and errors in a particular gravity condition, and then tested immediately afterward. At the completion of

a test in a given gravity condition, *Ss* moved on to the next gravity condition, as described by the schedule of counterbalanced orders, and were trained to asymptote and then tested. This procedure was carried out for all three gravity conditions. Time to perform each task, time between each task, and errors were recorded during each test. At the completion of the testing, *Ss* were asked to fill out questionnaires relating to performance in clothing and gravity conditions and performance of the various tasks.

RESULTS

The results reported in this section are divided into three parts: base-line training, specific training and test, and questionnaire responses.

Base-Line Training

The mean number of shirt-sleeve, 1-gravity training trials to performance asymptote was 8.89 with a standard deviation of 2.31 trials. The mean time to perform one torquing task element at asymptote was 11.37 sec. with a standard deviation of .81 sec. The mean time to perform one connector-mating task element at asymptote was 3.12 sec. with a standard deviation of .25 sec. The mean time to perform one threading task element at asymptote was 8.53 sec. with a standard deviation of .60 sec. It is important to remember that the performance times reported refer to mean time to perform one task element, for example, the mean time to torque one bolt, instead of the total time to perform all nine bolt torques. A total of one error was committed by the nine *Ss* at base-line training asymptote.

Specific Training and Test

All nine *Ss* received base-line training in the shirt-sleeve, 1-gravity mode. At the completion of this training, *Ss* received specific training on all tasks in a particular clothing condition and in a given gravity condition. Since the gravity conditions were counterbalanced, individual *Ss* required different numbers of training trials for different clothing conditions and different gravity sequences. The latter depended on the amount of transfer of training from base-line training. For example, training and test in 1 gravity had complete transfer to the 1-gravity-harnessed condition while transfer from 1 gravity to ⅙ gravity showed much less transfer, depending on preceding specific training and test sequences.

The results of this experiment are summarized in Table 1 and Figure 1 for performance times. In general, there was a significance in performance times between the three clothing conditions ($p < .01$). Further analysis of this finding revealed that performance in the vented suit required more time than in the shirt-sleeve mode ($p < .01$). Performance in the pressurized suit required significantly more time than the vented suit ($p < .01$) and, therefore, required significantly more time than the shirt-sleeve condition. In terms of percentage of time increase, the vented suit showed a 63.61% increase over the shirt-sleeve mode. The pressure-suit condition showed a 127.70% increase over the shirt-sleeve condition and a 39.16% increase over the vented suit.

The effect of the three gravity conditions was significant beyond the .01 level of confidence. An analysis of the specific components of this variable revealed no significant difference between the 1-gravity condition and the 1-gravity-harnessed condition. However, performance in the ⅙-gravity mode took significantly longer than in the 1-gravity-harnessed condition ($p < .01$), and significantly longer than in the 1-gravity condition. With respect to percentage of increase in time, the 1-gravity-harnessed condition showed a 1.8% increase over 1 gravity. The ⅙-gravity mode showed an increase in performance time of 23.52% over 1 gravity, and an increase of 21.33% over the 1-gravity-harnessed condition.

The difference in performance times for each of the three tasks was significant beyond the .01 level of confidence. Using a 1-degree-of-freedom test, it was demonstrated that the threading task took significantly longer than the connector-mating task ($p < .01$). Further, it was demonstrated that the torquing task took significantly more time to perform than the threading task ($p < .01$). In terms of percentage of time increase, the threading task showed a 192.24% increase over the connector-mating task. The torquing task showed a 267.18% increase over the con-

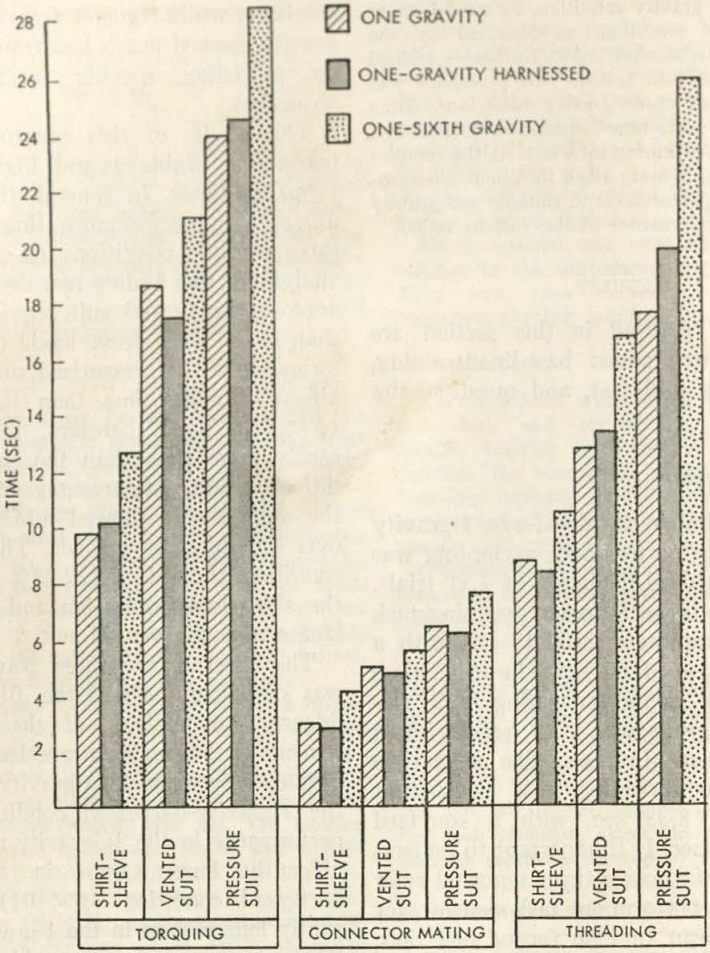


FIG. 1. Performance times by task for all gravity and clothing conditions.

nector-mating and a 25.64% increase over the threading task.

The interaction between clothing and gravity was not significant. This was interpreted to mean that for every level of the gravity condition, performance in the clothing condition remained relatively constant. In short, the ⅙-gravity condition required greater time to perform each of the tasks regardless of clothing condition, and regardless of clothing condition, 1 gravity and 1 gravity harnessed required about the same time to perform.

The interaction between clothing condition and task was significant ($p < .01$). This was interpreted to mean that the difficulty of the tasks did not remain constant for every level of clothing condition.

Specifically, the threading task when in the vented suit showed less of an increase in performance time over the shirt-sleeve condition than did the torquing task in the same clothing mode (56.51% versus 76.68%). However, a comparison of performance times in the pressurized-suit mode revealed that the tasks were approximately equivalent for this condition (130.55% versus 138.48%). Further, the connector-mating task showed much less of a performance-time increase in the pressure-suit mode (87.53%) than did either the threading (130.55%) or the torquing (138.48%) tasks.

TABLE 1

STATISTICAL ANALYSIS OF PERFORMANCE TIMES

Source	*SS*	*df*	*MS*	*F*
Clothing (A)	1,349.31	2	674.65	26.92**
Subjects within groups (B)	150.34	6	25.06	
Gravity (C)	128.83	2	64.42	26.40**
Task (D)	2,630.99	2	1,315.50	539.14**
A × C	16.64	4	4.16	1.70
A × D	349.69	4	87.42	35.83**
C × D	30.21	4	7.55	3.09*
B × C	41.28	12	3.44	1.41
B × D	100.39	12	8.36	3.43**
A × C × D	19.72	8	2.46	1.01
B × C × D	58.57	24	2.44	

* $p < .05$.
** $p < .01$.

The interaction between gravity and task was significant ($p < .05$). Again, the difficulty of the tasks did not remain constant for all gravity conditions. The finding is supported by the fact that the threading task in 1/6 gravity took 35% more time to perform than in 1 gravity.

The interaction between *S*s within groups and gravity condition was not significant. All *S*s experienced a decreased efficiency in 1/6 gravity while performing approximately the same in either 1 gravity or 1 gravity harnessed. However, the interaction between *S*s and task was significant ($p < .01$). Different tasks were of varying difficulty for each *S* within a given clothing condition. Finally, a comparison of the intertask intervals revealed no differences between task, clothing, or gravity conditions. The mean interval was 15 sec.

The term "error," as employed in this research, referred to the frequency of occurrence of dropping any task element, or holding on to any part of the performance panel or structure framework while performing any of the tasks. There was no attempt made to assign relative weights according to the nature of the error. The Friedman two-way analysis of variance by ranks revealed that performance as measured by errors did not differ significantly for any of the three main variables, and there were no significant interactions. These findings were due, in part, to the loose definition of "error."

Questionnaire Responses

Subjective impressions of test *S*s regarding the task under various gravity conditions were examined for conformity with the objective data reported above. All nine *S*s agreed that the harness, per se, made no discernible difference. Therefore, the primary differences of concern were between 1 gravity and 1/6 gravity for the various suited conditions.

In this comparison, whenever a number is reported in quotes, it refers to the scale number on a scale from "1" (easy) to "5" (very difficult), and is a mean representation of the three responses.

The *S*s were asked to rate the entire maintenance task from easy to very difficult. The shirt-sleeve *S*s rated the task as easy in both 1 and 1/6 gravities, "1" and "1.67," respectively. The vented-suit *S*s found the task moderately difficult in both gravities, "3" and "3.67." The pressure-suit *S*s rated task difficulty as "3" for both gravity conditions.

When asked to identify the most difficult single task, the connector-mating task was never indicated. All three shirt-sleeve *S*s considered the threading task most difficult in 1 gravity, and considered the torquing task most difficult in 1/6 gravity. Two vented-suit *S*s considered the threading task to be the most difficult in 1 gravity, and all three identified the torquing task as the most difficult when in 1/6 gravity. Torquing and threading were identified as equally difficult by the pressure-suited *S*s in both gravities.

The *S*s did, in no instance, identify the middle portion of the performance panel as difficult to work on. Two shirt-sleeve *S*s identified the lower portion as the most difficult to work on in both gravities. The higher portion of the panel was identified as the most difficult by all three vented-suit *S*s. However, two pressure-suit *S*s described the lower portion of the panel as most difficult to work on. In short, it appears that extreme positions, high or low, caused the greatest performance difficulty relative to maintenance-task operations.

The shirt-sleeve *S*s did not consider the maintenance task fatiguing in either 1 or 1/6 gravity. Their ratings were "1" and "1.67,"

respectively. The vented-suit *S*s found the task moderately fatiguing, "3," for both gravities. The pressure-suited *S*s rated the task as "3" in 1 gravity and "4" in ⅙ gravity. The indication here is a trend in clothing condition going from the shirt-sleeve to pressure-suit mode. It is reasonable to suggest that, if the pressure-suit *S*s were required to finish all task elements, they would have reported extreme fatigue.

The *S*s in all three clothing conditions unanimously agreed that they felt they had committed the most errors on the threading task, irrespective of gravity condition. They also unanimously agreed that they had to change their method of performance of the tasks when in ⅙ gravity as compared to performance in 1 gravity.

The vented-suit *S*s reported that performance dexterity was moderately reduced, "3," when compared to their shirt-sleeve base-line training. The pressure-suit *S*s rated suit encumbrance as "5," or extremely restrictive. Finally, no pressure-suit *S*s reported anxiety in working in the pressurized mode.

DISCUSSION

The results of this study supported the hypothesis that a maintenance-task decrement would be manifested in lunar gravity when compared with performance in the earth's gravity. The finding of no statistical difference between the 1-gravity and 1-gravity-harnessed conditions indicated that the performance decrement in lunar gravity was not due to the requisite harness. The *S*s' responses to the questionnaire also supported these findings: "In lunar gravity, I stood on tiptoes. More difficult to aim screws or nuts due to less stability. Torquing produced counter torques pushing me away from the panel."

Perhaps the most significant finding of this research was that *S*s performing in the pressurized-suit condition were unable to complete all of the tasks. Further, the results of an analysis of the performance times supported the hypothesis that the shirt-sleeve clothing condition would impose the least restriction on performance while the pressurized suit would impose the greatest performance restriction. These findings concurred with Holmes' (1965) description of the pressurized-suit mode as being the single most important variable hindering maintenance tool performance.

The results of this research also supported the hypothesis that the connector-mating task would take less time to perform, while the torquing task would require the greatest amount of time due to counter-torques. The hypothesis that the threading task would cause the greatest number of errors was also supported. The interactions involving task indicated that maintenance tasks may be affected differentially by the person performing them, the clothing condition, and the gravity condition in which they are performed.

Task Analysis and Timeline Studies

During successive stages of design and development of a complex system, information regarding the nature of the human interface with the system in terms of crew functions and tasks is necessary in order to insure that the system does not impose performance requirements beyond man's capabilities. One purpose of this research was to derive extrapolation figures for maintenance-type tasks so that performance-time estimates could be made more accurately for lunar maintenance timeline studies.

The performance data, summarized in Figure 1, provide an empirical basis for assigning performance times to related lunar mission activities. For example, in estimating performance times for lunar intravehicular tasks in shirt-sleeves, 25.67% increment in performance time should be made over comparable earth-based performance. Such generalization from the data obtained in this study would provide a more valid basis for performance-time allocations than the "armchair analysis" technique so often resorted to.

Performance Aids

The findings of this research, that man's behavior changes under the influence of lunar gravity, concur with the lunar-gravity research literature. The performance decrement observed in lunar gravity and in the various clothing conditions points to the necessity of considering human engineering design counter-

measures for coping with the peculiarities of the partial gravity lunar environment. While considerable attention is currently being given to performance aids for assisting the astronaut in the zero-gravity environment (Pierson & Geller, 1965), relatively little attention has been devoted to a comparable study of performance aids required to facilitate lunar maintenance and operational tasks (Holmes, 1965). Based on the results of the present research, it may be just as necessary for the astronaut to anchor himself to his task in the lunar environment as it is for earth orbital operations.

The following represents guidelines for unique design conceptualization of performance aids. Handholds should be provided for the astronaut primarily as a method for obtaining a satisfactory position relative to the task. However, if possible, the astronaut should not be required to use his hands in steadying his position relative to his equipment task since this would limit his ability to use both hands in accomplishing tasks. It appears that a one-hand grip would not be sufficient to steady his position for many tasks. Properly spaced and distributed toe holds and tether anchor points appear to offer promise in helping to steady the astronaut's position for tasks involving force application such as the torquing task. Holmes (1965) investigated tether concepts in lunar gravity for maintenance tool work and concluded that a single tether located from the navel area of the suit to the task was the best candidate. Further research is required. It has also been pointed out that the pressurized suit prohibited performance in extreme positions. In addition to the tethers and handholds just cited, extravehicular work may require lunar ladders or toe holds on the side of structures where work is required above the astronaut's shoulders. Also, work areas extremely close to the lunar surface should be avoided where possible.

REFERENCES

HAZARD, A. B. Results of preliminary physiological testing under simulated lunar and martian gravity conditions. *American Institute of Aeronautics and Astronautics*, 1965, 3, 296–302.

HEWES, D. E., & SPADY, A. A., JR. Evaluation of a gravity-simulation technique for studies of man's self-locomotion in lunar environment. NASA Tech. Note D-2176, 1964, 1–34. (a)

HEWES, D. E., & SPADY, A. A., JR. Moon operations here on earth. *Aeronautics and Aerospace Engineering*, 1964, 24–28. (b)

HOLMES, A. E. Space tool kit survey, development, and evaluation program final report. Martin Company Report ER-13942, 1965, 1–186.

PIERSON, W. R., & GELLER, R. E. Work in low friction environment. *American Institute of Aeronautics and Astronautics*, 1965, 3, 1074–1079.

PRESCOTT, E. J., & WORTZ, E. C. Metabolic costs of upper torso maneuvers under reduced-gravity conditions. *Aerospace Medicine*, 1966, 37, 1046–1049.

ROBERTS, J. F. Walking responses under lunar and low gravity conditions. Aerospace Medical Research Laboratory Tech. Report 63–112, 1963, 1–117.

(Received May 17, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 3, 184–187

MEMORY DYNAMICS AND WORK MOTIVATION[1]

H. MELTZER AND D. LUDWIG[2]

Washington University

Of 143 workers asked to report their pleasant and unpleasant experiences, only 24 included work as 1 of their outstanding pleasant experiences. These 24 composed the work-motivated group. A group equated for age and job level was selected as a control group. The 2 groups were compared on measures of: (a) memory optimism, determined by predominance of pleasantness over unpleasantness; (b) ratings of their adequacy as workers; (c) autonomy; (d) interpersonal competence. Significant differences were found in favor of the work-motivated group: they were more often memory optimists and superior in work adequacy, autonomy, and interpersonal competence. The relationship of these findings to the motivation-to-work controversy is considered.

About work motivation and satisfaction almost everything has been said, but not everything is known. Since Herzberg, Mausner, and Snyderman (1959) published their book, *The Motivation to Work,* in which they advocated a two-factor theory of job satisfaction, Herzberg and his supporters have published many articles to prove this theory. However, other contributors studying the same problem are not at all convinced that Herzberg is right. Evidence negative to Herzberg's theory has been reported in studies by Dunnette (1965), Ewen (1964), Friedlander (1963), and others. A review of the literature indicates that those studies which best support Herzberg's theory are those which use the same or similar methods. Quinn and Kahn (1967), in their review of organizational psychology, have this to say about the Herzberg controversy:

> Whereas Herzberg has rested a stimulating theory of job satisfaction on a rather shaky empirical foundation, the Cornell group (1963–65) have launched an intensive series of studies of the measurement of job satisfaction with only a minimum of theory to guide them [p. 456].

According to the Cornell researchers (Ewen, Smith, Hulin, & Locke, 1966), their conclusion is that neither the Herzberg theory nor the traditional theory are supported by the data they investigated. Instead, results indicate that intrinsic factors are more closely related to overall satisfaction and overall dissatisfaction than the extrinsic factor, pay, and suggest that functioning of the extrinsic variable may depend on the level of satisfaction with the intrinsic variables. In the literature on work motivation by Herzberg and protagonists, as well as by his antagonists, the problem of personality differences between a motivated group and a nonmotivated group has not been studied.

Vroom (1962) hypothesized and confirmed that if a person believes himself to possess an ability and believes that successful performance of his task requires that ability, he will prefer performing the task effectively to performing it ineffectively. If he believes effective performance is irrelevant to his ability, he will be indifferent about performing. If effective performance is negatively relevant, he will prefer to perform ineffectively.

Porter (1963a, 1963b), in a series of studies of managers, showed that their satisfactions were substantially related to such variables as level of job, line-staff type of position, and size of company. His studies are based on a modified form of Maslow's hierarchy of needs. In that sense it can be considered a problem of motivation and satisfaction. However, he, too, has neglected to study personality characteristics of work-motivated people.

In the first study on the problem of memory dynamics of workers, Meltzer and Ludwig (1967) found that, though younger workers are significantly more often memory optimists than older workers, each age group significantly expressed pleasant memories more often

[1] Based on a report presented at the Divison 14 session on Organizational Behavior at the American Psychological Association, Washington, D. C., September 9, 1967.

[2] Now at Concordia Senior College, Fort Wayne, Indiana.

than unpleasant. These findings were interpreted as suggesting that personality factors, rather than age factors as such, influence memory dynamics involved. In a study still unpublished, a relationship was found between memory dynamics of workers and certain personality variables. While working with the memories reported, it was apparent that only a few included work as one of their outstanding pleasant experiences of a lifetime. Of the 143 workers, only 24, or 16.78%, included work as one of their outstanding pleasant experiences. It dawned on the investigators that here, if ever, was a work-motivated group. Not only did work have intrinsic satisfaction to them, but work evidently was ego-involving enough to be included in pleasant experiences of a lifetime. It would be interesting to compare workers of that kind with those to whom work did not have that kind of ego-involving meaning. Accordingly, it was arranged to compare an equated group with them on four measures—memory optimism, work adequacy, autonomy, and interpersonal competence.

The purpose of the present study was to fill, in some small measure, the gap in knowledge about the relationship of work satisfaction to personality characteristics by comparing a work-motivated group with a non-work-motivated group in certain personality variables, including memory dynamics. The problems studied, expressed in the form of hypotheses, were: (1) More of the work-motivated group will be memory optimists; (2) the work-motivated group will be significantly higher in work adequacy than the non-work-motivated group; (3) the work-motivated group will be significantly superior in the two personality variables measured—autonomy and interpersonal competence.

METHOD

Subjects and Setting

The *S*s used in this study were 143 workers in a paper converting industry located in a stable college community in upstate New York. This paper mill is the only plant of any size in the community, and obtaining a job there is considered to be a sign of establishment to many people living in the community. For the most part, *S*s were semiskilled workers with large families and strong feelings about their jobs in the industry. The *S*s ranged in age from 19 to 78 and represented about one-third of all the workers in the plant. The 143 were selected from a larger group because they were the only ones interviewed for data concerning pleasant and unpleasant experiences. All of the workers who mentioned their work experiences as a part of their pleasant memories were classified in the work-motivated group. A group matched to these 24 workers was selected from the remaining 119 workers by equating them for age and job level.

Source Material

In the interview as structured, the directed questioning about work, family, personal story, and present conditions and problems was presented in open-end fashion. The more projective phases (Meltzer, 1950) of the interview, which in this study include the recall of outstanding pleasant and unpleasant experiences of a lifetime, were given after the factual phase of the interview was completed.

Procedure

The interview was structured in the following sequence:

1. Direct questioning about work, family, personal story, and present conditions or problems.
2. Recall of the most outstanding pleasant experience with the question, "And now tell me what comes to your mind when you think of the most outstanding pleasant experience of your life."
3. Recall of other outstanding pleasant experiences.
4. Recall of the most outstanding unpleasant experience.
5. Recall of other outstanding unpleasant experiences.

Memory optimism was determined by predominance of pleasantness over unpleasantness in the experiences given by the *S*; memory pessimism was determined by the predominance of unpleasantness over pleasantness. In the original article, however, wishful and eternal optimists were differentiated from realistic optimists (Meltzer & Ludwig, 1967).

Work competence was based on the rating by the interviewers at the end of the interview. The interviewers had available the ratings of the respective foremen of the *S*s studied to consider as an aid in making their judgments. Scores for various ratings were as follows: 1 for D; 2, D+; 3, C; 4, C+; 5, B; 6, B+; 7, A; 8, A+.

Autonomy was determined on the basis of a content analysis of the personal and work histories given by the *S*, including projective interview material. The *S* was given a "+" in autonomy if the histories contained statements indicating adequate security and adjustment; a "—" indicated insecurity and maladjustment; and a "0" was assigned if no clear assessment could be made. A second judge made independent ratings of the same factors, with a correlation of .6 between the two. (A 3×3 contingency table was used with the maximum correlation being .82.)

Interpersonal competence was also determined by a content analysis of the same source material, namely the work and family histories, plus projective responses to memories of pleasantness and unpleasantness. An *S* was given a "+" in interpersonal competence if the histories contained statements indicating that *S* got along well with others. If there were statements indicating a history of poor relations with others a "−" was given. A "0" was given if no clear judgment could be made. The rating of the second judge correlated .64 with the first. (A 3 × 3 contingency table was used with the maximum correlation being .82.)

RESULTS

Differences in Memory Optimism and Work Adequacy (Hypotheses 1 and 2)

The two groups were compared on memory optimism and work adequacy, or performance. The means, standard deviations, and *t* values for the two groups on memory optimism and work adequacy are reported in Table 1.

These results support the first two hypotheses; namely, that the work-motivated group is composed of people who are more likely to be memory optimists, and that the work-motivated group rates significantly higher on work adequacy.

Differences in Autonomy and Interpersonal Competence (Hypothesis 3)

The *S*s were then compared on the measures of personality variables—autonomy and interpersonal competence. A comparison was made in a 2 × 2 fashion with the chi-square values as a measure of significant difference. The relevant data for the comparison of the two groups are presented in Table 2. On the measure of autonomy, the work-motivated group rated significantly higher than the matched group. On the measure of interpersonal competence, the work-motivated group was rated more favorably than the matched group. These two findings support the third hypothesis; namely, that the work-motivated group is significantly superior on the two personality variables—autonomy and interpersonal competence.

TABLE 1

DIFFERENCES BETWEEN WORK-MOTIVATED GROUP AND CONTROL GROUP ON MEMORY OPTIMISM AND WORK ADEQUACY

Factor	Group	*M*	*SD*	*t* value
Memory optimism	Work-motivated	2.125	1.12	3.47*
	Non-work-motivated	.875	1.20	
Adequacy	Work-motivated	5.502	2.32	2.89*
	Non-work-motivated	3.461	2.11	

* $p < .01$.

TABLE 2

DIFFERENCES BETWEEN WORK-MOTIVATED GROUP AND CONTROL GROUP ON AUTONOMY AND INTERPERSONAL COMPETENCE

Factor	Group	Positive factor	Negative factor	χ^2
Autonomy	Work-motivated	11	3	9.03**
	Non-work-motivated	5	16	
Interpersonal competence	Work-motivated	14	6	4.04*
	Non-work-motivated	5	19	

* $p < .05$.
** $p < .01$.

DISCUSSION

Significant characteristics which differentiate the work-motivated group from the control group are revealed in this study. In every respect measured, the work-motivated group comes out more favorably or superior than the equated control group of fellow workers. The work-motivated group contains significantly more realistic memory optimists, who are rated more adequate as workers on the job, are more autonomous or secure and adjusted, and are more competent in their interpersonal relations. The two groups are greatly different in memory dynamics and job adequacy as well as personality variables. In this plant, at least, people who are work motivated as evidenced by their ego involvement in the form of expressed feelings in their life memories are identified by better performance, better adjustment, and more optimistic outlook on their past lives. The procedure described here has been and can be used to advantage for management-evaluation purposes. For research purposes it would be worthwhile to compare the findings based on a similar study with the results obtained on the same workers with the use of the Critical Incident Approach of Herzberg. Such exploration opens up problems of significance, the

study of which could help resolve some of the conflict about motivation to work found in the literature of today.

REFERENCES

Cornell studies on job satisfaction. Cornell University, Department of Psychology, Ithaca, New York, 1963–65.

DUNNETTE, M. D. Factor structures of unusually satisfying and unusually dissatisfying job situations for six occupational groups. Presented at the meeting of the Midwestern Psychological Association, Chicago, April 1965.

EWEN, R. B. Some determinants of job satisfaction: A study of the generality of Herzberg's theory. *Journal of Applied Psychology*, 1964, **48**, 161–163.

EWEN, R. B., SMITH, P., HULIN, C. L., & LOCKE, E. A. An empirical test of the Herzberg two-factor theory. *Journal of Applied Psychology*, 1966, **50**, 544–550.

FRIEDLANDER, F. Underlying sources of job satisfaction. *Journal of Applied Psychology*, 1963, **47**, 246–250.

HERZBERG, F., MAUSNER, G., & SNYDERMAN, B. *The motivation to work*. New York: Wiley, 1959.

MELTZER, H. Memory dynamics, projective tests, and projective interviewing. *Journal of Personality*, 1950, **19**, 48–63.

MELTZER, H., & LUDWIG, D. Age differences in memory optimism and pessimism in workers. *Journal of Genetic Psychology*, 1967, **110**, 17–30.

PORTER, L. W. Job attitudes in management: III. Perceived deficiencies in need fulfillment as a function of line versus staff type of job. *Journal of Applied Psychology*, 1963, **47**, 141–148. (a)

PORTER, L. W. Job attitudes in management: IV. Perceived deficiencies in need fulfillment as a function of size of company. *Journal of Applied Psychology*, 1963, **47**, 386–397. (b)

QUINN, R. P., & KAHN, R. L. Organizational psychology. In P. R. Farnsworth (Ed.), *Annual review of psychology*. Vol. 18. Palo Alto, Calif.: Annual Reviews, Inc., 1967. Pp. 437–466.

VROOM, V. H. Ego-involvement, job satisfaction, and job performance. *Personnel Psychology*, 1962, **15**, 159–177.

(Early publication received November 17, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 3, 188–194

SUPERVISORY INFLUENCE AND WORK-GROUP PERFORMANCE

KURT R. STUDENT

Center for Research on Utilization of Scientific Knowledge, University of Michigan

Incremental influence, a new concept for the study of organizational leadership, was empirically tested by relating 12 work-group performance measures with measures of supervisors' social power. Referent power and expert power were conceptualized as incremental influence and considered qualitatively different from reward power, coercive power, and legitimate power. Referent power and expert power are considered to be idiosyncratic in character and depend upon an individual's unique role behavior, while reward power, coercive power, and legitimate power are organizationally determined and designed to be equal for all supervisors on the same hierarchical level. As predicted, expert power was related to a low accident rate, referent power was somewhat related to measures of subordinate withdrawal, and both referent power and expert power were generally related to group measures of production performance. These results are regarded as supporting the concept of incremental influence.

A recent book by Katz and Kahn (1966) has suggested a new concept for the study of organizational leadership. The concept, incremental influence, is stated as follows:

> When we think of leadership in contrast to routine role performance, however, we become particularly interested in those kinds of individual behavior which go beyond required performance and realize more fully the potential of a given position for organizational influence. In other words, *we consider the essence of organizational leadership to be the influential increment over and above the mechanical compliance with routine directives of the organization* [p. 302].

The present research is an empirical investigation of the concept of incremental influence in an industrial organization. More specifically, this research will analyze some performance correlates of the first-line supervisors' incremental influence.

A five-fold typology of social power suggested by French and Raven (1959) was used as the framework for the author's analysis of supervisory influence. Although other bases of power could be cited, French and Raven specify these five as the primary by which an agent, O, can exert influence over a person, P: (1) *Reward power,* based on P's perception that O has the ability to mediate rewards for him. In order for O to use reward power to influence P, he must possess resources that P values, and P must believe that by conforming he will actually benefit by the resources in question. (2) *Coercive power,* based on P's perception that O has the ability to mediate punishments for him. Analogous to the situation for reward power, in order for O to use coercive power to influence P, P must want to prevent some act of O and believe that through conforming he will do so. (3) *Legitimate power,* based on internalized values in P that O has a right to influence P and that P has an obligation to accept this influence. (4) *Referent power,* based on P's identification with O. French and Raven define identification as "a feeling of oneness of P with O, or a desire for such identity." If O is a person toward whom P is attracted, P will have a desire to be closely associated with O. This identification of P with O can be established or maintained if P behaves, believes, or perceives as O does. (5) *Expert power,* based on P's perception that O has some special knowledge or expertness in a given area. French and Raven hold that the strength of expert power varies with the extent of knowledge which P attributes to O and that the range of expert power is restricted to those cognitive areas in which O is believed to have superior knowledge. Experience, training, reputation, demonstrated ability, etc. are among the many reasons why P attributes expertness to O.

This five-fold typology of the bases of supervisory power makes possible an important conceptual distinction between the bases of power and leads to an operationalization of the concept of incremental influence. The areas in which reward power, coercive power, and legitimate power may be exercised are largely specified by the organization; on the other hand, areas in which the supervisor can exercise referent power and expert power are to a substantial degree uniquely determined by his own behavior and his interactions with his subordinates (expert power is personally determined, but the amount of information available may vary as a part of organizational structure). The extent and range of a supervisor's referent power and expert power cannot be specified by the organization. Expert power and referent power are idiosyncratic in character. On the other hand, reward power, coercive power, and legitimate power are nomothetic in character and result from the supervisor's occupancy of a position in the organization's role system. The formal organizational structure is designed to provide equal legitimate power to all supervisors on the same hierarchical level and to give them equal access to the use of organizational rewards and punishments. However, supervisors are not equal in their referent power or their expert power; the supervisor's referent power and his expert power constitute his particular utilization of his formal position. Thus the author concludes that referent power and expert power are conceptually and qualitatively different from reward power, coercive power, and legitimate power. Influential acts based on referent power and expert power constitute an increment in organizational influence beyond the influence inherent in the routine functioning of the organization's role system. Consequently, referent power and expert power are the basis of incremental influence, and incremental influence has been operationalized as the additive combination of referent power and expert power.

The present research relates incremental influence to work-group performance. Since recent research in a variety of organizational settings casts doubt on the validity of job performance as a unidimensional construct, a multiple-criteria approach to work-group performance was used. Supervisory influence will be related separately to 12 performance measures. These measures were conceptualized by a framework suggested by March and Simon (1958) who state that organizational members face two different decisions—the decision to participate and the decision to produce. Consequently, the performance measures are considered as measures of withdrawal, or as measures of production; accident rates are considered separately since this criteria does not fit into the withdrawal-production classification. An impressive body of research evidence (Kahn & Katz, 1960; Likert, 1961; Mann, 1965; Pfiffner, 1955) suggests that incremental influence (i.e., both referent power and expert power together) would relate to production measures. On the other hand, March and Simon (1958) hypothesized that withdrawal measures relate to satisfaction but not to performance, which in the author's terms would suggest that withdrawal measures relate mainly to referent power. Two studies support this view. Fleishman, Harris, and Burtt (1955) found a negative relationship between supervisory consideration and absenteeism; Fleishman and Harris (1962) found a negative relationship between supervisory consideration and turnover. Finally, in terms of the analytical model, the author considers that accidents are unintended consequences resulting from low supervisor "technical competence" (Mann, 1965, defines technical competence as the ability to use pertinent knowledge, methods, techniques, and equipment necessary for the performance of specific tasks and activities) and the inability of the supervisor to impart his competence to his subordinates. Accordingly, one would expect that accidents relate mainly to supervisory expert power. Consequently, the operational hypotheses were that incremental influence (i.e., both referent power and expert power together) would relate to the production measures; referent power would relate to the withdrawal measures; expert power would relate to the accident rate.

TABLE 1

RELATIONSHIPS BETWEEN MEASURES OF SUPERVISORY INFLUENCE

Measure	1	2	3	4	5	6
1. Incremental influence						
2. Referent power	.83***					
3. Expert power	.89***	.49**				
4. Reward power	.44**	.32*	.43**			
5. Coercive power	.12	.12	.09	.45**		
6. Legitimate power	.35*	.38*	.24	.08	.41**	

Note.—Group means, $N = 40$.
* $p < .05$, two-tailed.
** $p < .01$, two-tailed.
*** $p < .001$, two-tailed

METHOD

Research Site and Sample

This study employed data collected from a manufacturer of major home appliances. The data were collected for a larger study sponsored by the company and conducted by the Survey Research Center. The sample consisted of 486 hourly employees and 39 first-line supervisors representing 40 of the 52 work groups which comprised the company's two main plants, and the unit of analysis in this research was the work group. Measurements of attitudes, opinions, and perceptions of behavior were collected by paper-and-pencil questionnaires administered to nonsupervisory production employees during March 1964.

Questionnaire Measures of Supervisory Influence

The measures of supervisory influence were as follows:

Referent power. A single-item rating by subordinates of the extent to which they comply with their supervisor's directives because he is a "nice guy" and they don't want to hurt him.

Expert power. A single-item rating of the extent to which subordinates comply with their supervisor's directives because they respect his experience and good judgment.

Reward power. A single-item rating of the extent to which subordinates comply with their supervisor's directives because he can give special help and benefits to those who cooperate with him.

Coercive power. A single-item rating of the extent to which subordinates comply with their supervisor's directives because he can penalize or make things difficult for those who do not cooperate with him.

Legitimate power. A single-item rating of the extent to which subordinates comply with their supervisor's directives because he has a right, considering his position, to expect subordinates to do what he wants.

Incremental influence. As noted earlier, this concept was operationalized by summing the group mean for referent power and the group mean for expert power.

Table 1 presents the relationships between these measures of supervisory influence, while Table 2 presents the means and standard deviations of incremental influence and the bases of social power.

Measures from Company Records

For purposes of the larger study from which the present research data are taken, a roster of work-group performance measures was developed so that a number of different measures of performance could be related to questionnaire data. Twelve of these work-group performance measures were used in the present research. These 12 measures were tabulated weekly by the company during a 17-week period from immediately preceding and following the questionnaire administration. Six of the measures were direct measures of frequency or amount, and these measures were converted to a per-employee basis for each work group; the other six measures were based on ratings of each work group by the department head. The ratings were on 4-point descriptive scales based on the firm's existing budgetary, allowance, or schedule standards. Actual group performance was compared against

TABLE 2

MEAN AND STANDARD DEVIATION OF INCREMENTAL INFLUENCE AND BASES OF POWER

Measure	*M*	*SD*
Incremental influence	6.12	.92
Referent power	2.73	.47
Expert power	3.39	.59
Reward power	2.78	.40
Coercive power	2.32	.42
Legitimate power	3.43	.44

Note.—Group means, $N = 40$.

TABLE 3

LISTING AND DESCRIPTION OF WORK-GROUP PERFORMANCE MEASURES

Rating measures of performance
Indirect costs (+)
Performance rating based on actual versus budgeted amounts of inspection, stockhandling, general labor, make-up allowances, and rework labor (4-point scale).
Maintenance costs (+)
Performance rating based on actual versus budgeted amounts for maintenance, considering both type and quality of maintenance orders (4-point scale).
Supply costs (+)
Performance rating based on actual versus budgeted allowance for supplies (4-point scale).
Scrap costs (+)
Performance rating based on actual versus budgeted amounts for scrap and rework (4-point scale).
Performance against schedule (+)
Performance rating based on the extent to which the work group: stayed on or ahead of schedule (Rating = 4), required help to stay on schedule (Rating = 3), was behind schedule (Rating = 2), or was behind to the extent of causing down-time (Rating = 1).
Quality (+)
Performance rating based on actual versus anticipated numbers of rejects, taking into consideration the disposition, amount, and kind of reject (4-point scale).
Average earnings (+)
The average earnings per hour per man for each work group.
Excused absences (−)
The number of occurrences of excused absence, divided by the number of men in the work group.
Unexcused absences (−)
The number of occurrences of unexcused absence, divided by the number of men in the work group.
Accidents (−)
The number of reported injuries divided by the number of men in the work group.
Turnover (−)
The number of quits, transfers, and formal bids for jobs in other work groups and not involving advancement, divided by the number of men in the work group.
Suggestions submitted (+)
The number of suggestions formally submitted, divided by the number of men in the work group.

Note.—In each case, direction of scale is indicated by a plus or minus sign. "+" means that a high numerical rating represents "good" performance.

these abstract standards. Table 3 lists the 12 work-group performance measures taken from the company records and presents a description of each measure.

Note that a high score on any 1 of the 12 work-group performance measures may indicate either excellent or poor performance depending upon the specific measure in question.

RESULTS

The first hypothesis predicted that expert power would be positively related to a low work-group accident rate. The results from an analysis of this prediction are presented in Table 4 indicating that the hypothesis is supported. Expert power is the only one of the five bases of supervisors' power to relate significantly to a low accident rate.

The second hypothesis predicted that measures of withdrawal—excused absences, unexcused absences, and turnover—would be related to referent power. Specifically, the author predicted that high referent power would be associated with low absences and low turnover. The results relating to this prediction are presented in Table 5.

Table 5 indicates that this prediction is only partially supported. As expected, referent power is associated with low excused absences. However, it was found that unexcused absences and turnover are not associated as expected with supervisory bases of power. Surprisingly, high referent power tends to be related to *high* turnover. This latter finding is contrary to most previous research which would suggest that supervisory behavior which is characterized by high referent power should result in employees' continuing in the system. In the present case, however, unexcused absences are not at all related either to referent power or to the other bases of supervisory power, while turnover is related in a direction opposite to that predicted.

The third hypothesis predicted high incremental influence (both referent power and expert power) would be related to good performance on those measures which were considered to represent measures of production. The results from an analysis of this prediction are presented in Table 6.

The predictions of this study are generally supported. For four production performance measures (indirect costs, supply costs, quality,

TABLE 4

RELATIONSHIPS BETWEEN ACCIDENTS AND SUPERVISORY POWER

	Subordinate work-group's mean perception of their supervisor's					
	Incremental influence	Referent power	Expert power	Reward power	Coercive power	Legitimate power
Accidents	−.24	−.12	−.28*	−.03	−.16	−.20

Note.—Group means, $N = 40$.
* $p < .05$, one-tailed.

TABLE 5

RELATIONSHIPS BETWEEN WITHDRAWAL PERFORMANCE MEASURES AND SUPERVISORY POWER

Withdrawal performance measure	Subordinate work-group's mean perception of their supervisor's					
	Incremental influence	Referent power	Expert power	Reward power	Coercive power	Legitimate power
Excused absences	−.36*	−.35*	−.28*	−.18	.16	−.12
Unexcused absences	.00	−.02	.02	.18	.02	−.08
Turnover	.12	.23	−.01	.14	.08	.01

Note.—Group means, $N = 40$.
* $p < .05$, one-tailed.

and suggestions submitted) the predicted zero-order relationships are significant. For two other production performance measures (scrap costs and maintenance costs) the predicted zero-order relationships tend toward significance. However, no relationships exist for two predicted relationships; incremental influence is not related to good performance on average earnings or performance against schedule.

DISCUSSION

Although legitimate power ranked first among the reasons for compliance with supervisory directives, legitimate power is not related to differences in the performance of

TABLE 6

RELATIONSHIPS BETWEEN PRODUCTION PERFORMANCE MEASURES AND SUPERVISORY POWER

Production performance measure	Subordinate work-group's mean perception of their supervisor's					
	Incremental influence	Referent power	Expert power	Reward power	Coercive power	Legitimate power
Indirect cost performance	.27*	.40**	.10	.15	.22	.00
Maintenance cost performance[a]	.12	.00	.18	−.20	−.30*	.10
Supply cost performance	.31*	.21	32*	.31*	.08	.08
Scrap cost performance	.25	.33*	13	.26	.12	.06
Performance vs. schedule	−.11	.05	−.21	−.06	.04	−.05
Quality	.36*	.32*	.31*	.13	−.08	.11
Average earnings	.01	.00	.01	−.40**	−.22	.05
Suggestions submitted	.28*	.36*	.14	.09	.40**	.10

Note.—Group means, $N = 40$.
[a] Mean for 29 groups used as measure for mean of remaining 11 groups.
* $p < .05$, one-tailed.
** $p < .01$, one-tailed.

these subordinate work groups. The data suggest, therefore, that when the supervisor has a choice in the bases of power he emphasizes, to achieve compliance with his directives the supervisor should attempt to create and to use referent power or expert power rather than rely too strongly on his legitimate power.

Reward power and coercive power, unlike legitimate power, do relate to a number of performance measures. However, the overall relationship between these two bases of power is two-sided. From the organization's point of view the correlates of reward power and coercive power are problematic and have unintended consequences for the organization. Organizationally speaking, reward power relates functionally to supply costs but dysfunctionally to average earnings; coercive power relates functionally to suggestions submitted but dysfunctionally to maintenance costs. Therefore the data suggest that the supervisor's use of reward power and coercive power should be tempered where possible by the creation of referent power and expert power. Since there appear to be no dysfunctional or organizationally unintended correlates of incremental influence, the latter should be created and used where possible.

As was noted above and as predicted, incremental influence has positive and organizationally functional correlates. However, it is important to note that only 7 of the 12 performance measures relate to incremental influence. Five measures, two withdrawal measures (unexcused absences and turnover) and three production measures (maintenance costs, performance against schedule, and average earnings), showed no relationship to incremental influence. The reason for the difference in relationship between incremental influence and performance is not altogether clear. Perhaps the five performance measures which do not relate to performance represent areas or factors over which the first-line supervisor has little or no control. Unexcused absences and turnover might be related to extraorganizational considerations or to organizational factors which are unrelated to compliance with the first-line supervisor. The three production measures which do not relate to incremental influence may be said to represent company-wide or system factors rather than to be related to factors at the first-line supervisory level of the company. Maintenance costs, performance against schedule, and average earnings may be related to company-wide policies regarding maintenance programs, production schedules, and pay rates, respectively.

Finally, it should be noted that the correlations for referent power and expert power are not generally additive, and referent power is a better predictor of performance than is expert power. Referent power is significantly related to five performance measures (excused absences, indirect costs, quality, suggestions submitted, and scrap costs), while expert power relates significantly to four measures (accident rate, excused absences, supply costs, and quality). Only in the case of quality is the correlation for incremental influence greater than the correlations for referent power and expert power, which suggests that in this case the bases of power are additive.

The present research has shown the concept of incremental influence to be a useful and viable one. The qualitative distinction between referent power and expert power on the one hand and reward power, coercive power, and legitimate power on the other was generally supported by the data. The performance correlates of referent power and expert power are surprisingly large in view of the fact that the research did not include extraorganizational factors and included only one level of management. Moreover, the first-line supervisor in the present research site is a supervisor who may be said to have a relatively small area of freedom; his role range is limited by industrial engineers, quality-control specialists, time and motion study men, and others, as well as by the standardized nature of high-volume, single-product production. In view of the limitations inherent in the role which was studied, the findings are encouraging and further research and validation of these findings are indicated.

REFERENCES

FLEISHMAN, E. A., & HARRIS, E. F. Patterns of leadership behavior related to employee grievances and turnover. *Personnel Psychology*, 1962, **15**, 43–56.

FLEISHMAN, E. A., HARRIS, E. F., & BURTT, H. E. *Leadership and supervision in industry*. Columbus, Ohio: Bureau of Educational Research, Ohio State University, 1955.

FRENCH, J. R. P., JR., & RAVEN, B. H. The bases of social power. In D. Cartwright (Ed.), *Studies in social power*. Ann Arbor, Mich.: Institute for Social Research, 1959. Pp. 150–167.

KAHN, R. L., & KATZ, D. Leadership practices in relation to productivity and morale. In D. Cartwright & A. Zander (Eds.), *Group dynamics*. (2nd ed.) Evanston, Ill.: Row, Peterson, 1960. Pp. 554–570.

KATZ, D., & KAHN, R. L. *The social psychology of organizations*. New York: Wiley, 1966.

LIKERT, R. *New patterns of management*. New York: McGraw-Hill, 1961.

MANN, F. C. Toward an understanding of the leadership role in formal organization. In R. Dubin, G. C. Homans, F. C. Mann, & D. C. Miller (Eds.), *Leadership and productivity*. San Francisco: Chandler, 1965. Pp. 68–103.

MARCH, J. G., & SIMON, H. A. *Organizations*. New York: Wiley, 1958.

PFIFFNER, J. M. The effective supervisor: An organization research study. *Personnel*, 1955, 31, 530–540.

(Received May 25, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 3, 195–199

INFLUENCE OF TRAINING, METHOD, AND RELATIONSHIP ON THE HALO EFFECT

EVA METZGER BROWN [1]

Teachers College, Columbia University

120 student nurses each rated 6 peers on 10-point scales for 6 sets of traits. The variance of a judge's ratings for each object was found, and the variance scores of the 6 objects for each judge summed. This dispersion score was used as the measure of the halo effect. It was found that trained judges exhibited less halo than untrained judges, but that ratings made by rating 1 individual on 6 traits at a time yielded substantially the same amount of halo as did ratings made by rating all individuals on 1 trait at a time. The relationship between the judge and the object was a significant variable for the untrained judges. Here the ratings for the less intense relationship exhibited significantly more halo than the ratings for the intense relationship. Training of the judges improved all ratings to the point where the judge-object relationship did not play a significant role.

The halo effect poses a problem for the investigator because of its pervasive yet elusive presence in the rating situation. It occurs when "a particular rater tends to rate a particular ratee similarly on all traits" (Guilford, 1959, p. 146), for any reason. The halo effect was noted briefly by Wells in 1907 and labeled by Thorndike in 1920.

Statistically, the halo effect has been defined in three ways: (*a*) as an inflated intercorrelation among traits (Gilinsky, 1947; Taylor & Hastman, 1956), (*b*) as a general bias factor arrived at through matrix and factor analyses (Grant, 1952; Meyers, 1965), and (*c*) as the judge-object (*J–O*) interaction error evaluated according to Guilford's (1954) analysis-of-variance model. The first definition, although responsible for the original concept of the halo effect, is now viewed as lacking in sophistication. The findings of factor-analytic studies, while of value in identifying halo, are limited to a specific experimental situation. The third definition is the most sophisticated in that it differentiates between objective (valid) and relative (invalid) halo, and the more useful in that it is applicable to all studies where a number of *J*s are rating common *O*s on a common scale. Although the data of the present study were not based on ratings of common *O*s, the conception of Guilford (1959) was used to define halo as the *J–O* interaction error and to apply the analysis-of-variance technique. In the study, the halo effect was defined as inversely proportional to the variance score for ratings given an *O* by a *J* over six sets of traits.

The main purpose of this study was to identify the conditions which minimize the operation of the halo effect. The study examined the influences of the training of the *J*s, of the method of rating, and of the quality of the *J–O* relationship. Previous research has indicated that *J*s benefit from training when it includes: (*a*) practice with the specific rating scale (Wakeley, 1961), (*b*) discussion of errors in rating by the *J*s (Levine & Butler, 1952), and (*c*) special emphasis on the importance of trait differentiation (Taylor & Hastman, 1956). As for the method of rating, Symonds (1925) suggested that more halo results when a *J* rates one individual at a time on all traits (Method 1), than when a *J* rates all individuals on one trait at a time (Method 2). Method 2 has been widely accepted as superior to Method 1, but recent studies (Blumberg, DeSoto, & Kuethe, 1966; Johnson, 1963; Taylor & Hastman, 1956) have reported substantially the same degree of halo effect for both methods. Characteristics of the *J–O* relation-

[1] The research reported in this paper is based on a doctoral dissertation submitted to Teachers College, Columbia University. Special thanks are given to Laurance F. Shaffer, chairman of the dissertation committee, Elizabeth P. Hagen and Rosalea A. Schonbar, members of the committee, and to Rosedith Sitgreaves, consultant to the project.

ship which influence the *J*'s ratings of the *O* have not been studied adequately. Greater attention has been focused on the level of acquaintance between the *J* and the *O* (Bare, 1954; Koltuv, 1962), and the data support the hypothesis that increased acquaintance is related to decreased bias in judgment.

The broad hypothesis of the study is that the operation of the halo effect is influenced by the training of the *J*s, by the method of rating used by the *J*s, and by the intensity of the relationship between the *J* and the *O*.

METHOD

Subjects

One hundred and twenty student nurses in attendance at St. Luke's Hospital School of Nursing, New York, served as subjects. They were divided into four groups: A1, A2, B1, and B2. Groups A were trained, whereas Groups B were not trained; Groups 1 made ratings according to Method 1, whereas Groups 2 made ratings according to Method 2.

Instruments

The instruments were designed and adapted to serve the needs of the study, and written instructions preceded each questionnaire.

The Information Questionnaire was used to gather some background information on each *J*. The Form for the Selection of Objects required each *J* to select three *O*s toward whom she felt as favorable as possible (intense relationship) and three *O*s toward whom she felt indifferent (less intense relationship). All *O*s were fellow students.

The Acquaintance Checklist assessed how well the *J* felt she knew the *O*, while the Social Distance Scale measured how well the *J* liked the *O*.

The Rating Scale consisted of six bipolar traits (factors) on 10-point scales. The factors, selected from the Guilford-Zimmerman Temperament Survey (Guilford & Zimmerman, 1949), included lethargic–active, impulsive–self-controlled, withdrawn–sociable, emotionally unstable–emotionally stable, mannish–feminine, and hostile–friendly.

The Postexperimental Reaction Questionnaire tapped the *J*'s knowledge of the aims of the study and her previous experience in rating others.

Procedure

The data were collected in four 1-hr. periods spaced 1 wk. apart. No more than 30 *J*s were seen per session and no more than 15 *J*s were trained at a time.

In Session I a broad outline of the procedure relevant to the trained and untrained groups was given, and the confidentiality of ratings was stressed. The *J*s completed the Information Questionnaire and selected their *O*s.

In Session II, Group A was trained, while Group B did not meet. The experimenter, acting as an instructor, told Group A about different kinds of rating scales and rating procedures, problems in obtaining sound ratings, and constant errors. The *J*s were given some practice on the Rating Scale and their ratings were discussed.

In Session III, the *J*s completed the Acquaintance Checklist, the Social Distance Scale, and the Rating Scale. In Session IV, these three scales were readministered and their test-retest reliabilities found to be .95, .96, and .85, respectively. Finally, the Postexperimental Questionnaire was completed.

RESULTS

Characteristics of the J–O Relationship

Student nurses were selected as subjects because there was a basic level of familiarity among them. In the sample used, however, 91% of the "favorable" *O*s were known "extremely well" or "very well" by the *J*s, whereas 62% of the "indifferent" *O*s were known only "a little" or "hardly." The correlation between the Acquaintance Checklist and Social Distance Scale was .79. It was concluded that the *O*s who had an intense relationship with the *J* were well known to the *J*, and the *O*s who had a less intense relationship with the *J* were less well known to the *J*.

Influence of Training and Method on the Halo Effect

The means and standard deviations for the variance scores of the experimental groups appear in Table 1. Each mean value was based on 30 scores, except for the combined "Groups" and "Methods" where the mean value was based on 60 scores. Each score was the sum of one *J*'s variance scores for three favorable or three indifferent *O*s, or the sum of one *J*'s variance scores for six *O*s when the relationships were combined. The higher the score the smaller was the halo effect, and the lower the score the greater was the halo effect.

The influence of training and method on the halo effect was evaluated by a two-way analysis of variance (Table 2). The significance of the variable of training at the .01 level led to the conclusion that a brief but specific training period leads to more analytic and discriminating judgments.

The variable of method was nonsignificant,

TABLE 1

MEANS AND STANDARD DEVIATIONS OF RATING-SCALE VARIANCE SCORES ACCORDING TO TRAINING, METHOD, AND RELATIONSHIP

Experimental groups	Method 1	Method 2	Both methods
Trained			
Intense relationship			
$\bar{X}$	9.95	9.53	9.74
s	6.12	7.21	6.64
Less intense relationship			
$\bar{X}$	13.08	8.33	10.71
s	7.30	3.44	6.14
Both relationships			
$\bar{X}$	23.03	17.86	20.44
s	11.24	9.58	10.68
Untrained			
Intense relationship			
$\bar{X}$	7.16	9.23	8.19
s	4.50	4.75	4.71
Less intense relationship			
$\bar{X}$	7.04	6.36	6.70
s	4.56	5.31	4.92
Both relationships			
$\bar{X}$	14.20	15.59	14.89
s	7.72	8.17	7.91
Both groups			
$\bar{X}$	18.61	16.73	17.67
s	10.57	8.90	9.77

indicating that ratings made according to Method 1 and Method 2 did not differ substantially from one another. A significant interaction between training and method at the .05 level was found to exist. A further evaluation of differences using Scheffé's method indicated that the trained and untrained groups differed significantly only when they made ratings according to Method 1. Hence, the significant interaction reflected the effect of training on Method 1.

TABLE 2

TWO-WAY ANALYSIS OF VARIANCE: THE INFLUENCE OF TRAINING AND METHOD ON THE HALO EFFECT

Source of variation	df	SS	MS	F
Training	1	935.36	935.36	10.86[b]
Method	1	103.50	103.50	1.20
Interaction	1	341.32	341.32	3.96[a]
Within groups	116	9985.31	86.08	
Total	119	11365.49		

[a] $F_{.95} = 3.92$.
[b] $F_{.99} = 6.85$.

Influence of the I–O Relationship on the Halo Effect

The influence of the *I–O* relationship on the halo effect was tested by chi-square. The basic score was the ratio of the sum of the variance scores for the three high-intensity *Os*, over the sum of the variance scores for the three low-intensity *Os*. Each *I* was represented by one value or ratio score. The null hypothesis of the chi-square test was that, with an *F* distribution of 15 and 15 degrees of freedom, the 30 values (30 per group) are randomly observed. The degrees of freedom were based on the number of *Os* per relationship (3) multiplied by one less than the number of traits (5). The probability tables for such an *F* distribution are presented by Guenther (1965, p. 298). As the smaller the variance score the greater the evidence for the halo effect, it followed that when the ratio value was less than one, the ratings of the favorable *Os* exhibited greater halo than those ratings of the indifferent *Os*. When the ratio value exceeded one, the ratings of the indifferent *Os* exhibited greater halo than those ratings of the favorable *Os*.

The chi-square test for each of the four experimental groups indicated that there was no significant difference between Method 1 and Method 2 for both the trained and untrained groups, so that the data were combined (Table 3). The null hypothesis of the chi-square was accepted for the trained group and rejected for the untrained group, at the .05 level of significance. A difference between

TABLE 3

CHI-SQUARE: THE INFLUENCE OF INTENSITY OF RELATIONSHIP ON THE HALO EFFECT WITH TRAINING HELD CONSTANT

Quarters	F distribution of ratio scores	Expected frequency	Observed frequency	
			Trained	Untrained
1	.000– .701	15	21	13
2	.702–1.000	15	11	11
3	1.001–1.430	15	14	10
4	Over 1.430	15	14	26
	Chi-square		3.6	11.1[a]

[a] $\chi^2_{.95} = 7.8$.

the intense and less intense relationship was thus found only for the untrained group. An examination of the mean variance scores (see Table 1) indicated that for the untrained group the ratings of the less intense relationship exhibited a greater amount of halo, while for the trained group the ratings became differentiated to the point where the relationship between the *J* and the *O* did not play a significant role.

DISCUSSION

The conception of the measure of halo as a variance score was a new one. The variance score was a direct statistical translation of the definition of halo which stated that "a particular rater tends to rate a particular ratee similarly on all traits" (Guilford, 1959, p. 146). With the variability of ratings built into the basic score, individual differences in the relative amount of halo among *J*s could be identified. In addition, group differences could be found through the analysis-of-variance technique so that conditions which functioned to increase or decrease the amount of halo in ratings could be identified. This procedure is practical and can be applied to any setting where *J*s are rating *O*s on graphic rating scales. However, a shortcoming of the procedure is that the statistical significance of the objective and relative components of the halo effect cannot be determined independently as in Guilford's (1954) model. One has to assume that a large difference in variance scores indicates significant variation in the relative component of the halo effect. Such a difference was found between the trained and untrained groups.

Training small groups of inexperienced *J*s was very effective in reducing the amount of halo even when such training was brief (1 hr.). The fact that the subjects were accustomed to the role of the student and accepted the experimenter as an instructor probably had a positive influence on the findings.

The training process was affected by two uncontrolled features, the first of which led to a difference in the observed motivational level between the trained and untrained groups. It is generally assumed that *J*s who are more involved and more interested in the task will make more careful and more accurate ratings (Conrad, 1932). In the study, the trained group had more contact with the experimenter and more general information about the study. It was observed that the trained *J*s wrote more on the Reaction Questionnaire and exhibited greater interest in the results of the study than the untrained group. Therefore, it is proposed that the trained *J*s had more invested in the ratings they were making than the untrained *J*s and hence took more care in making discriminating ratings. Further exploration of other effects of the training period, apart from learning, is required. For example, to what degree is the amount of halo in ratings affected by increased experimenter-*J* contact or by a desire to please the experimenter?

The second uncontrolled feature of the training process lay in the content of the training session and is proposed as an explanation for the significant interaction found between training and Method 1. During the training session, practice in rating and discussion of ratings were based solely on Method 1. It follows that *J*s who rated according to the same method by which they were trained could apply what they had learned more directly than could the *J*s who rated according to Method 2 and were trained according to Method 1. This shortcoming in the study points out the importance of having the training session equally relevant for all the variables in the experiment.

As for the method of rating, a sizable sample of student nurses rated *O*s with whom they were familiar similarly whether they used Method 1 or Method 2. Added to the accumulated body of research (Blumberg et al., 1966; Johnson, 1963; Taylor & Hastman, 1956), the findings affirm the irrelevance of this methodological variation for the halo effect. Therefore the often suggested recommendation that ratings should always be made by Method 2 (Guilford, 1954, p. 279) seems poorly founded in the case of graphic rating scales.

The significance of the *J–O* relationship for the halo effect varied with the training or lack of training of the *J*s. For the untrained group, the ratings of indifferent *O*s, the less intense relationship, exhibited greater halo

than the ratings of favorable *Os*, the intense relationship. As the indifferent *Os* were less well known than the favorable *Os*, the findings of Bare (1954) and Koltuv (1962), that increased acquaintance is related to decreased bias in judgment, seem to be supported. However, in the present study none of the *Os* was as "unfamiliar" to the *Js* as was the case in Koltuv's study, nor as briefly known as was the case in Bare's study. The student nurses had been going to school and living together for at least a year and a half, so that it could be assumed that the *Js* had some basis on which to make differentiated ratings for the indifferent as well as the favorable *Os*. A possible explanation of the findings was offered by the *Js*, themselves. On the Reaction Questionnaire, they reported difficulty in selecting *Os* for the less intense relationship. Apparently, there was greater resistance to and anxiety in selecting indifferent *Os* than favorable *Os*, for both the trained and untrained group. As the *J–O* relationship did not play a significant role in the ratings of the trained group, it is proposed that the training period not only functioned as a learning experience, but also tended to reduce the anxiety of the *Js* and hence the restrictiveness of their ratings. The latter function would be particularly pertinent for ratings of *Os* toward whom the *Js* felt "less favorable" and possibly ambivalent. Certainly, the variable of the *J–O* relationship needs to be explored further. How is the amount of halo affected when the level of acquaintance and the social distance are varied with each other? Does the amount of halo increase with increased social distance or is there a U-shaped distribution where the amount of halo is greatest for ratings of *Os* toward whom the *J* is ambivalent?

The questions posed by this study reveal the need for a further understanding of the halo effect and the conditions which affect it. Examined as unitary factors, the variables of training and relationship were found to have a significant influence on the amount of halo exhibited, whereas the variable of method was nonsignificant. It follows that the conditions of training and *J–O* relationship need to be analyzed into their significant components, and the contribution of these characteristics to the halo error assessed.

REFERENCES

BARE, R. H. Bias as related to rater contacts. Unpublished doctoral dissertation, Ohio State University, 1954.

BLUMBERG, H. H., DESOTO, C. B., & KUETHE, J. L. Evaluation of rating scale format. *Personnel Psychology*, 1966, **19**, 243–259.

CONRAD, H. S. The personal equation in ratings. I. An experimental determination. *Pedagogical Seminary and Journal of Genetic Psychology*, 1932, **41**, 267–297.

GILINSKY, A. S. The influence of the procedure of judging on the halo effect. *American Psychologist*, 1947, **2**, 309–310.

GRANT, D. L. An exploratory study of halo effect in ratings. Unpublished doctoral dissertation, Ohio State University, 1952.

GUENTHER, W. *Concepts of statistical inference*. New York: McGraw-Hill, 1965.

GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.

GUILFORD, J. P. *Personality*. New York: McGraw-Hill, 1959.

GUILFORD, J. P., & ZIMMERMAN, W. S. *Guilford-Zimmerman Temperament Survey Manual*. Beverly Hills, Calif.: Sheridan Supply Co., 1949.

JOHNSON, D. M. Reanalysis of experimental halo effects. *Journal of Applied Psychology*, 1963, **47**, 46–47.

KOLTUV, B. B. Some characteristics of intrajudge trait intercorrelations. *Psychological Monographs*, 1962, **76**(33, Whole No. 552).

LEVINE, J., & BUTLER, J. Lecture vs. group decision in changing behavior. *Journal of Applied Psychology*, 1952, **36**, 29–33.

MEYERS, J. H. Removing halo from job evaluation factor structure. *Journal of Applied Psychology*, 1965, **49**, 217–221.

SYMONDS, P. M. Notes on rating. *Journal of Applied Psychology*, 1925, **9**, 188–195.

TAYLOR, E. K., & HASTMAN, R. Relation of format and administration to the characteristics of graphic rating scales. *Personnel Psychology*, 1956, **9**, 181–206.

THORNDIKE, E. L. A constant error in psychological ratings. *Journal of Applied Psychology*, 1920, **4**, 25–29.

WAKELEY, J. H. The effects of specific training on accuracy in judging others. Unpublished doctoral dissertation, Michigan State University, 1961.

WELLS, F. L. A statistical study of literary merit. *Archives of Psychology, New York*, 1907, **1**, No. 7.

(Received May 26, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 3, 200–204

DRIVER GAP ACCEPTANCE AT INTERSECTIONS

WILLIAM L. GIBBS

United States Bureau of Public Roads, Washington, D. C.

To determine the most probably correct theory of the several currently accepted theories of driver gap acceptance, the gap limen was investigated by the method of constant stimuli. In a field-test situation, drivers were exposed to an approaching vehicle for a 1-sec. interval at various times during its approach. Approach speeds of 16, 32, 48, and 64 mph were used. Analysis of the data indicates drivers accept a gap equal to 5.4 times the approach-vehicle velocity (expressed in ft/sec). The obtained function most nearly supports the time hypothesis of driver gap acceptance.

Driver decisions at street intersections require a complex series of judgments. The driver is required to detect any approaching vehicle, determine range and speed of approach, then decide whether to cross, or merge with, the traffic stream. Although drivers accomplish the task with great frequency and a high degree of safety, comparatively little is known about the actual process or processes utilized by the driver. Early experimenters (*E*s) made gross measurements of vehicle movements, and developed theories of gap acceptance based on these measurements, but comparatively little is known about the actual decision processes of the driver at the intersection.

Development of several different theoretical approaches to the driver's acceptance of gaps at intersections has followed the initial study by Greenshields (1947). Early *E*s measured and cataloged the gap acceptance of drivers in free traffic situations to develop the early expressions of the model of driver gap acceptance. Following are the major hypotheses relating gap acceptance to various aspects of cue structure or cognitive organization of such cues.

1. Time hypothesis: The Greenshields model considered gap acceptance judgments as if the driver made a precise estimate of arrival time of the approaching vehicle. Thus a driver would make the judgment that there was some time period before the oncoming car would arrive at the intersection and thus he could, or could not, cross, or merge with the mainstream traffic.

2. Modified-time hypothesis: Evolving from the time hypothesis, a more complex expression for the time-gap acceptance decision process was developed by R. L. Moore (1956). Working with pedestrians, he required them to make an estimate of the last possible moment they would attempt to cross a roadway. (They did not actually cross.) The function he developed was of the order $D = 45 + 2.9V$, where $D =$ distance in feet from the pedestrian to the approaching vehicle, and $V =$ approach speed of the vehicle in feet per second. His experiment covered the range of approaching vehicle speeds of 15–60 mph with 100 observations at each data point. One particular item of interest is that as speed of the approaching vehicle (V) increased, there were more potentially hazardous errors in that insufficient time was allowed to cross the street. In fact, at 60 mph there were 35 judgments where insufficient time was allowed for crossing of the 100 total judgments made at that speed.

3. Distance hypothesis: A third model, that of constant distance, has been postulated. This theoretical approach says that the driver will accept a gap of a constant distance regardless of the speed of the approaching vehicle.

Brain (1962) in a largely theoretical treatment of gap acceptance considered several additional models:

4. Rate of change of angular position: The apparent movement of the approaching vehicle in relation to the coordinates of the head of the driver in space. Brain showed that these cues are of limited value because the rate of change is approximately zero when the two cars are approaching each other at speeds

which would produce a collision, and also when the driver's car is halted at the stop line.

5. Rate of change of horizontal dimension due to apparent rotation: As the vehicle approaches, the changing appearance of the car as it changes from head-on to more and more side view may provide the cues to cross or not to cross. Brain contends that the values derived from the condition of two cars on a collision course with the driver's vehicle halted at the stop line will provide insufficient cues. This would, in addition, require the postulation of a mechanism to develop the length/breadth ratio required and then integrate the length/breadth ratio over time to develop the decision. Brain considers that the complexity of such mechanism makes this model highly improbable.

6. Rate of change of dimension due to distance: Brain worked with height of the vehicle to derive a theory of angular velocity similar to that of Michaels and Weingarten (1965). Although Brain's (1962) theoretical derivation indicates this model is most descriptive of the process the driver follows, after experimental work he states, "It would appear . . . that one relies on visual image size based on past experience rather than calculus for estimating minimum time intervals [p. 791]."

7. Experience: The method, according to Brain, which is apparently used by drivers. The crossing decision is based on known size of the visual image and the expected performance of the approaching vehicle, and not on subtle cues derived from the change or rate of change of the image of the approaching vehicle.

For the most part, all preceding studies (which used *S*s) have used free observation wherein *S*s could see the approaching vehicle until he made his decision. The decision of *S* was thus a summation of information over an extended period of time.

In the present study, an information acquisition time (IAT) of 1.0 sec. was selected to permit a driver sufficient time to make a crossing decision. Use of an IAT of 1.0 sec. was considered to have two advantages: (1) it appeared to be close to the actual decision times utilized by drivers, and (2) it permitted a precise determination of the threshold of crossing decisions due to the experimental control over the cues used to make the gap acceptance decision.

This experiment was for the purpose of determination of the threshold of a driver's gap acceptance by the use of the psychophysical method of constant stimuli in order to permit a comparison of the different models of crossing behavior which have been proposed by several different *E*s.

METHOD

Subjects

Four male Bureau of Public Roads employees who had not had prior contact with the experiment acted as *S*s. Vision was tested and found to be within normal tolerances for all *S*s. Age ranged from 31 to 62 yr. with an average of 51 yr. Driving experience was commensurate with age.

Apparatus

Two automobiles were used. One was parked at an "intersection" while the other was the stimulus vehicle (SV).

The *S* was seated in the parked vehicle with his vision restricted by an electrically operated shutter mounted in the driver's side window of the automobile. A chin rest associated with the shutter assured precise positioning of *S*'s head. A system of pneumatic tubes and electronic timers controlled the shutter opening time and duration of 1.0 sec. Shutter opening time was controlled for each run by the preset times which were manually inserted into the timer before each run. Total time for each run was also recorded in order to measure accuracy of SV approach velocity.

Procedure

The testing was conducted on an airport runway 200 ft. wide and 3000 ft. long.

With one vehicle "stopped" at a simulated intersection with a through street, the SV approached from the left and passed in front of the *S* in the stopped vehicle. The centerline of the SV path was approximately 20 ft. from *S* and perpendicular to the centerline of the stopped vehicle.

Four approach speeds were used for the SV: 16, 32, 48, and 64 mph. Preset times were so determined that the approach times, the time between the closing of the shutter and the SV arriving at a point directly in front of *S*, were in the same increments for all speeds. Seven approach times were used ranging from 1.8 to 9.0 sec. in 1.2-sec. steps.

With 4 speeds and 7 approach times there were thus 28 conditions (cells). Although the method of constant stimuli was used, a maximum of 10 trials was presented for any cell due to time and cost of each trial. To further reduce the number of trials the method of sequential analysis was used. Necessary data for the sequential analysis computations were obtained from pilot data. Using cumulative totals

for each cell, further trials were discontinued when the number of "yes" or "no" responses exceeded the limits of the sequential analysis table.

The task of the *S* was to observe the approaching vehicle and make judgments. The following instructions were read to *S*:

> We are conducting a controlled study of people's ability to determine whether they have appropriate gaps in traffic so they can cross through intersections. Through our simulated intersection you are asked to observe the vehicle that is approaching just as though you were sitting at a stop sign at an intersection. At a certain period of time during the approach of the vehicle a shutter in this viewing device will open. You are to judge during this period whether or not you would accept the gap and cross through the intersection. Base your decision on the position of the car just as the shutter closes. Try to make judgments as you normally do when driving. Prior to the time when you are requested to make a judgment I will say "ready." At that time you should put your chin in the chin rest. Shortly afterward you will see the shutter open. After the shutter closes you are to indicate by saying "yes" if you would have crossed through the intersection, or "no" if you would not. Between each of these trials you may take your chin away from the chin rest and relax.

Intertrial interval was about 1.5 min. Each *S* received about 100 trials per day with rest periods after each 25 trials. During testing, *S*s read books or magazines between each trial.

RESULTS

In order to determine whether there was a difference between *S*s, a comparison was made of the number of acceptance judgments at each of the four speeds by use of the chi-square statistic. The results indicate the performance of *S*s was different: at 16 and 32 mph the difference was significant at the $p < .02$, while at 48 and 64 mph the difference was $p < .01$. Due to the significant differences obtained and the small number of *S*s, extreme caution should be used in interpretation of further analysis. Whenever possible, each *S* will be considered separately.

Determination of the limen and standard deviation of gap acceptance was performed for each *S* by the normal graphic process (see Table 1). The limen in the present experiment is the point at which *S* would theoretically accept a given gap on 50% of the trials. The obtained limen was the same as the approach time each man would accept 50% of the time. Although there was marked individual variation in threshold, individual slopes were similar, and the mean approach time accepted was remarkably constant over the speed range.

TABLE 1

TIME LIMEN AS A FUNCTION OF APPROACH-VEHICLE VELOCITY

Subject	Velocity of approach vehicle[a]			
	16	32	48	64
S_1				
Limen	3.05[b]	2.97	4.0	3.5
SD	.32	.32	.77	.54
S_2				
Limen	7.15	5.33	6.30	5.28
SD	.63	.32	.83	.31
S_3				
Limen	7.67	5.33	4.89	4.18
SD	.87	.33	.45	.33
S_4				
Limen	4.85	6.75	6.69	8.05
SD	.98	1.67	.75	.91
Mean limen	5.68	5.10	5.47	5.25
SD	.70	.66	.70	.52

[a] In miles per hour.
[b] In seconds.

Using the time gap obtained for each man, the gap distances which he accepted were calculated. These data, presented in Table 2, were corrected to reflect the precise velocity of the stimulus vehicle.

To permit comparison of the four currently accepted theories of gap acceptance, a plot was made of the curves derived from previously derived empirical and theoretical

TABLE 2

DISTANCE LIMEN AS A FUNCTION OF APPROACH-VEHICLE VELOCITY

Subject	Velocity of approach vehicle[a]			
	16	32	48	64
S_1	73.11[b]	141.55	289.64	336.56
S_2	171.38	254.03	456.18	507.72
S_3	183.85	254.03	354.08	401.95
S_4	116.25	321.71	484.42	774.09
M	136.15	243.07	396.08	504.84
SE	25.63	37.33	45.19	96.36

[a] In miles per hour.
[b] In feet.

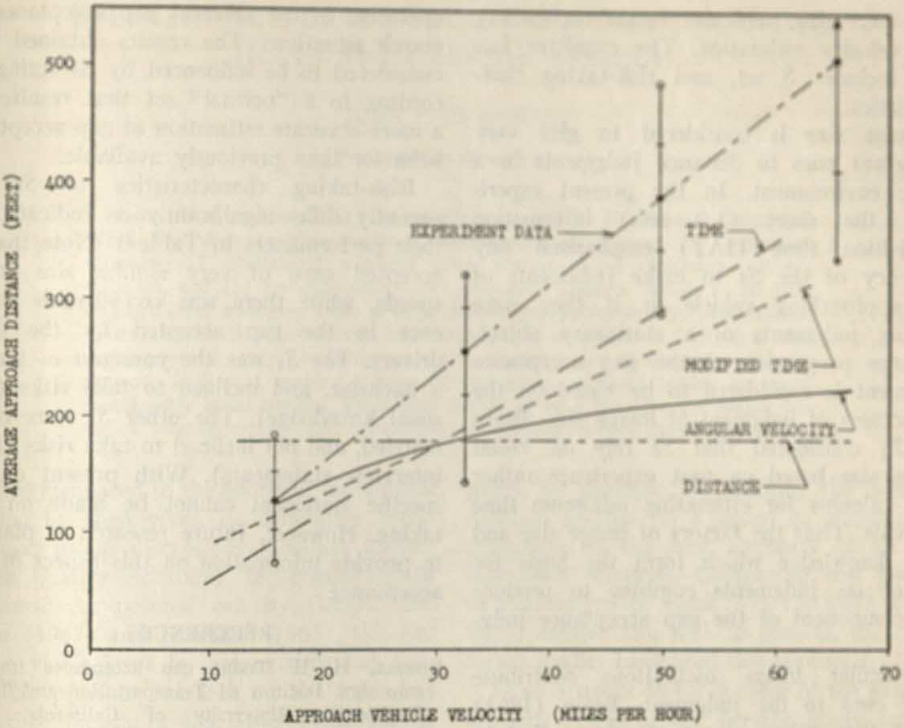


FIG. 1. Gap accepted as a function of approach-vehicle velocity.

values obtained from these theories (Figure 1).

The plot of the gap accepted versus the velocity of the approach vehicle permits graphic comparison of expected results which would be predicted by each theory. Data from the present experiment (Table 2) are included with the range of obtained values indicated by vertical lines and the crossbar indicating the standard error. These data produce a function which fit the equation $G = 5.4\ V$, where V = vehicle velocity in feet per second.

DISCUSSION

The approximately 5-sec. median gap size accepted by drivers in the present study agrees with Raff (1950) and Bissell (1960), and is only slightly shorter than that found by Robinson (1951) and Wagner (1966). In general, lower speed ranges were utilized in previous studies, but the congruence of the data, within similar speed ranges, indicates that the experimental controls imposed in the present experiment did not bias the results. The congruence of the data with most median accepted gaps in the 5–6 sec. range provides a strong indication that the time hypothesis is the most probable basis for driver gap acceptance decision.

Graphical analysis of the data also tends to support the time hypothesis as the basis for driver gap acceptance. Analysis of the data indicates the time hypothesis most closely approximates the correct description of the actual gap acceptance process. The results obtained in the present experiment strongly disagree with the modified-time hypothesis. Although at the two lower speeds it would appear that any of the hypotheses could be supported, the overall trend of the obtained function refutes these three hypotheses.

The constant necessary to approximate the fit between the time hypothesis and the obtained data requires some discussion. The marked change in gap distance required by *Ss* is considered to be a function of (*a*) perceptual processes and (*b*) cognitive factors. Perceptual factors include such factors

as: image size, vehicular image oscillations, and velocity estimation. The cognitive factors include: *S* set, and risk-taking characteristics.

Image size is considered to give very important cues to distance judgments in a static environment. In the present experiment the short (1.0 sec.) information acquisition time (IAT) emphasized any tendency of the *S*s to make judgments of the approaching vehicle as if they were making judgments of a stationary object. A large proportion of the gap acceptance judgment is considered to be based on the correctness of judgment of image size. Brain (1962) considered that *S*s rely on visual image size based on past experience rather than calculus for estimating minimum time intervals. Thus the factors of image size and prior knowledge which form the basis for image size judgments combine to produce one component of the gap acceptance judgment.

Vehicular image oscillations contribute some cues to the judgment. Brain (1962) stated that, "there was a conscious effort to allow for the speed of faster vehicles from the rate of apparent image growth (over the long viewing time he used) and from the amount of yaw, roll, and general visual response to carriageway irregularities . . . [p. 791]." Thus it is possible that *S*s in the present experiment did perceive slight vehicular oscillations during the short IAT and hence could make some judgments of relative approach speed. Using the oscillation cue as a modifier of the information from the image size/distance cue, *S* made a gap acceptance, or rejection, judgment.

The perceptual factors are further modified by cognitive factors, such as the set of the *S*. The present experiment required *S*s to respond "normally," whereas most other studies have requested *S* to judge "the last moment for crossing" (Brain, 1962; Moore, 1956). If *S*s were following the instructions, then the magnitude difference between the obtained data and previous data could be reconciled. However, the slopes obtained by previous investigators and the slope of the present data do not agree. Therefore some alternative, or additional, process may be operating in the different gap acceptance research situations. The results obtained were considered to be influenced by *S*s acting according to a "normal" set that resulted in a more accurate estimation of gap acceptance behavior than previously available.

Risk-taking characteristics of *S*s apparently differ significantly as indicated by their performances in Table 1. Note that S_1 accepted gaps of very similar size at all speeds, while there was considerable difference in the gap accepted by the other drivers. The S_1 was the youngest of the *S*s, a bachelor, and inclined to take risks (personal knowledge). The other *S*s were older, married, and not inclined to take risks (from interview statements). With present data a specific statement cannot be made on risk taking. However, future research is planned to provide information on this aspect of gap acceptance.

REFERENCES

BISSELL, H. H. Traffic gap acceptance from a stop sign. Institute of Transportation and Traffic Engineering, University of California, May 1960, Berkeley, California.

BRAIN, R. Vehicle speed estimation at right-angled at-grade junction, *Surveyor and Municipal and County Engineer,* 1962, 789–791.

CRAWFORD, A. The overtaking driver. *Ergonomics,* 1963, 6, 153–170.

GREENSHIELDS, B. D., SHAPIRO, D., & ERICKSEN, E. L. Traffic performance at urban street intersections. Tech. Report No. 1, Yale Bureau of Highway Traffic, Yale University, 1947.

MICHAELS, R. M., & WEINGARTEN, H. Driver judgment in gap acceptance. Office of Research and Development, United States Bureau of Public Roads, April 1965.

MOORE, R. L. Psychological factors of importance in traffic engineering. Paper presented during International Study Week in Traffic Engineering, Stresa, Italy, 1956.

RAFF, M. S. A volume warrant for urban stop signs. ENO Foundation for Highway Traffic Control, Sagatuck, Connecticut, 1950. Also available from Bureau of Public Roads Library, Wash., D. C.

ROBINSON, C. C. Pedestrian interval acceptance. *Proceedings of the Institute of Traffic Engineers,* 22nd meeting, 1951, 144–150.

WAGNER, F. A. An evaluation of fundamental driver decisions and reactions at an intersection. Highway Research Record No. 118. National Academy of Sciences, Highway Research Board, Wash., D. C., 1966. Pp. 68–84.

(Received May 29, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 3, 205–210

INDIVIDUAL VERSUS GROUP PROBLEM SOLVING IN AN INDUSTRIAL SAMPLE

JOHN P. CAMPBELL

University of Minnesota

A comparison of the quality of solutions to N. R. F. Maier's Change of Work Procedure problem under 3 experimental conditions: (1) individual solutions, (2) individual solutions after hearing group discussion without criticism, and (3) group solution after discussion, using a sample of 80 2nd- and 3rd-level managers from a large midwestern public utility. Motivational (Bass's Orientation Inventory) and biographical correlates of individual and group problem-solving behavior were also examined. Group discussion did not improve individual problem-solving efforts; group solutions tended to be of significantly lower quality. There also tended to be a negative correlation between solution quality and orientation toward maintaining group interaction.

When compared to the efforts of individuals working by themselves, do committee meetings and group discussions facilitate finding higher quality solutions to organizational problems? As pointed out by Gerard and Miller (1967) and Hoffman (1965), this has been a viable issue for a long time but considerable previous research on the topic has yielded conflicting results. Much of the earlier work tended to show that groups were better problem solvers than individuals (Lorge, Fox, Davitz, & Brenner, 1958), while more recent studies, such as those by Tuckman and Lorge (1962), Taylor, Berry, and Block (1958), and Dunnette, Campbell, and Jaastad (1963), have shown that under certain conditions a group may actually inhibit the best effort of its members. However, another recent study (Hall, Mouton, & Blake, 1963) tends to favor the superiority of the group solution.

The difficulties involved in interpreting much of the past research have been cited by Lorge et al. who point out that such investigations have frequently been carried out on ad hoc groups (in which the group members have not previously met and probably never will again) of students, and the experimental problems are often of the puzzle or anagram type. Another difficulty lies in the definition of what is an individual. Should the average group score be compared to the average individual score or to some index derived from the "nominal" group (Faust, 1959)? In a nominal group, individual efforts are combined in some fashion by an independent party, for example, by choosing the best individual solution in a group or obtaining a composite of the best elements of all the individual solutions. Comparing real groups to nominal groups seems to be the more meaningful comparison, and, in general, the more recent studies have used this approach.

In an attempt to fill in some of the gaps in previous research, the present study makes use of an industrial sample, a complex realistic problem, and compares the solution quality of real groups versus nominal groups as well as real groups versus individuals. In addition, an effort is made to relate the differential performance of an individual (isolation versus group discussion) to certain measures of individual differences. Attempts to find out what kinds of individuals are benefited or inhibited most by the group experience are totally lacking in the literature.

METHOD

Subjects

The *Ss* were 80 second- and third-line managers from a large midwestern public utility employing about 5,000 people. Most of the participants knew one another fairly well and many had previously been in meetings together. Thus, the groups used in the present study at least approach what Lorge et al. call a *traditional* group.

The Experimental Problem

The problem used was a slight modification of the Change of Work Problem developed by Maier and described in Maier, Solem, and Maier (1957). In its present form it reads as follows:

In a large firm which manufactures communication equipment, a four man department has the responsibility for the entire production of a particular component. Three of the men perform the assembly work and the fourth acts as a foreman. The assembly operation is divided into three positions and the workers have adopted a system of hourly rotation among the three jobs. The three jobs are not entirely routine and require somewhat different skills. However, all three of the men have been able to master each of the jobs at least minimally well but still perform some jobs better than others.

The foreman called a meeting to discuss the possibility of their changing their work method to one where each man works on only one position. This would be his best position according to the time study data given to the foreman. Although the new method should increase the productivity of the workers and thus their piece rate wages, the foreman's suggestion of a change to the new method was met with some resistance. The employees were concerned about the amount of boredom resulting from working at the same job all the time and believed that this might actually lower production. They were also a bit suspicious of management's motives in suggesting the new work method.

Can you suggest any solution to the problem faced by the foreman and his crew?

The problem has previously been scored by classifying solutions as old, new, and integrative. An old solution is simply to leave things as they are, and the new solution consists of adopting the work method proposed by management with maximum effort directed at communicating the mutual benefits of the work change to the employees. An integrative solution is a suggestion that the three men still rotate on the three jobs but that each man spend the most time on the job he does best, the second largest block of time on the job he does second best, and the least amount of time on the job he does most poorly. Thus, the integrative solution goes beyond selecting either the management or employee solution and suggests a unique alternative that combines the best elements of both. For the present study an attempt was made to develop a scoring system with a larger number of categories. The problem was first pretested on 43 individuals attending an extension course in supervision at the University of Minnesota and 38 individuals in a senior level course in personnel psychology. Most of the people in the extension class had supervisory responsibilities of some kind. These data were used to develop a wide range of possible solutions to the problem. Points were assigned to the alternatives roughly corresponding to Maier's progression from old to new to integrative. Scores could range from 0 to 60. Doing nothing received a score of 0. The alternative with the highest possible score entailed: (*a*) ranking jobs for each person in terms of how well he can perform each one; (*b*) having each person rotate among the three jobs such that he spends the most time on his best job, etc.; (*c*) carefully discussing the advantages and disadvantages of the new method in a group meeting with full opportunity for the employees to air their fears; (*d*) providing for the experimental tryout of various rotation schedules so that the employees and foreman will be better equipped to choose the best one; and (*e*) discussing the possibility of setting some production goals to shoot for during the transition period. Some representative alternatives between these two extremes are: rotating between the *two* most suitable jobs and spending more time on the more suitable (35 points); equal time rotation between the two best jobs but for periods longer than an hour (20 points); accepting management's new method without benefit of experimental tryout or group discussion of the pros and cons (5 points). At all the steps along the scale, bonus points are given for adopting an experimental outlook with regard to the various elements in the solution, providing for maximum two-way communication between management and the employees, and for attempting to set transitional production goals in consultation with the employees.

Other Measures

All *S*s in the study filled out an extensive 60-item personal-history questionnaire dealing with family and educational background, leisure-time activities, work history, participation in community affairs, etc., and also the Orientation Inventory (Bass, 1962) which yields three subscores: (1) self-orientation—which reflects the extent to which a person describes himself as expecting direct rewards to himself regardless of the job he is doing or the way he affects others; (2) interaction orientation—which reflects an individual's degree of concern with maintaining harmonious relationships in a superficial sort of way at the expense of concern for the task; (3) task orientation—which reflects the degree of concern with completing the job, solving problems, working persistently, and doing the best job one can. The test consists of 27 items all of which are keyed for each of the three scales.

Procedure

Forty-eight of the *S*s were randomly assigned to the experimental group and 32 to the control group. Initially all *S*s were mailed the biographical questionnaire and the personality measure. In the same mailing the experimental group was given the Change of Work Problem and asked to write what they considered to be the best solution, in as much detail as possible. The control group worked on the Circular Assembly Problem (Maier et al., 1957) which is also a combination production-process-human-relations type problem. Materials were originally mailed to 82 *S*s, and 81 were returned.

The *S*s were then assembled into 4-man groups, 12 experimental groups and 8 controls. All groups

were asked to work on the Change of Work Problem in a two-stage sequence. First the members of each group were asked to discuss and offer solutions to the problem using a brainstorming format; that is, no criticism or evaluation was allowed. They were just to explore as many solutions as possible. No time limit was imposed, but most groups seemed to be satisfied that they had covered everything after 30-60 min. of discussion. At this point *Ss* were asked to individually write down what they then considered to be the best solution to the problem. After these "individual-after-group-discussion" effects were collected, each group was asked to arrive at a consensus; one member of the group was asked to record the agreed-upon solution. This usually entailed another 20–40 min. of discussion.

Thus, for the experimental group three different solutions were offered to the problem: (1) an individual solution while working alone, (2) an individual solution after hearing and participating in a group discussion, and (3) a group solution. The control group offered solutions to the Change of Work Problem only under the last two conditions.

Assuming that the comparison of controls and experimentals does not show a significant practice effect for the experimental group, a number of different analyses are possible. Using individuals as degrees of freedom, comparisons can be made to see if group discussion increases or decreases the quality of individual solutions to the problem. Using real and nominal groups as degrees of freedom, comparisons can be made across all three experimental conditions. For the nominal groups (independent combinations of individual solutions) two different group scores were used, the average individual score in the group and a composite score obtained by putting together the best elements for each of the four solutions. The individuals in each nominal group were the same individuals who later made up a real group.

Scoring

The reliability of the scoring system was determined first on the pretest data by having three individuals (the writer and two graduate students)[1] independently score the 81 solutions and by computing the intercorrelations among the three scorers. It was checked again for two of the scorers on data from the experimental group. These two sets of scores were correlated for the 48 individual solutions to the Change of Work Problem and for the 48 individual-after-group-discussion solutions. Both scorers also derived and scored composite solutions from the 12 nominal groups and scored the consensus solutions from the 12 real groups.

The scores actually used for the experimental analysis were obtained by reconciling any differences between Scorer 1 and Scorer 2 to their mutual satisfaction.

[1] The writer wishes to thank Phillip Pettman and Martin Rock for their assistance as raters.

RESULTS

In general, the interscorer correlations were relatively high, ranging from .73 to .96 with an average of .82. Since the scores actually used in the analysis were the consensus of Scorers 1 and 2, the reliability of the composite ratings is probably higher, but it is impossible to estimate how much. The Spearman-Brown, for example, is inappropriate in this case.

For all types of scores, none of the differences between experimental and control groups was significant at $p = .10$ using a t test. Thus, a practice effect does not seem to be a crucial concern. Also, none of the treatment effects was significant when individuals were used as degrees of freedom; that is, there was no significant change in the average individual solution score after *Ss* participated in a brainstorming type group discussion.

For the group scores, differences across experimental treatments and the interactions of experimental treatments with the variables measured by the Orientation Inventory were examined by a series of two-way analyses of variance. Two levels of the orientation dimensions were obtained by dichotomizing scores at the median. The analyses for nominal and real groups were carried out for two different types of nominal group scores, the composite

TABLE 1

ANALYSES OF VARIANCE FOR GROUPS USING INTERACTION ORIENTATION SCORES AS LEVELS

Source	Composite score used as nominal group score			Average score used as nominal group score	
	df	*MS*	*F*	*MS*	*F*
Between					
Levels	1	393.4	2.40	50.1	1.99
Error	10	164.0		25.2	
Within					
Treatment	2	756.8	16.12*	62.6	10.12*
Treatment × Levels	2	35.2	.75	12.8	2.07
Error	20	46.9		6.2	

* $p = .01$.

TABLE 2

COMPARISONS OF MEAN GROUP SCORES CLASSIFIED BY EXPERIMENTAL CONDITION AND BY INTERACTION ORIENTATION SCORES

Orientation test scores	Mailed responses	After discussion	Group consensus	Combined
Composite score used as nominal group score				
High interaction orientation score	15.2	10.2	2.5	9.3
Low interaction orientation score	23.5	19.0	5.2	15.9
Combined	19.4	14.6	3.9	
Average score used as nominal group score				
High interaction orientation score	8.0	5.2	2.5	5.2
Low interaction orientation score	8.2	9.2	5.2	7.5
Combined	8.1	7.2	3.9	

score and the average of individual scores within the group.

The superiority of the nominal group was significant and large, regardless of which type of score was used to reflect the nominal group's performance. However, none of the main effects pertaining to the orientation variables and none of the interactions of these variables with the experimental treatments was statistically significant. Since the interaction orientation score was the only dimension to approach significance, only the analysis of variance for Experimental Treatment × Interaction is shown in Table 1. The corresponding means are shown in Table 2. The means and sums of squares for the treatments would of course be the same no matter which orientation score was used.

Correlations between the solution scores obtained under particular experimental treatments and the orientation data were also examined. For the individual scores the relationships were computed using the Pearson r, and for the analysis of group scores Spearman's rank-order correlation was used. Individual orientation scores were correlated with individual solutions obtained before discussion and after discussion, and with the difference. The rankings of average group orientation scores were correlated with the ranking for nominal group composite scores obtained before discussion, the ranking for the group consensus scores, and the difference between these two rankings. For example, a positive correlation between an orientation score and a difference score means that the higher an individual or group stands on the orientation variable the greater the decrement in solution quality under the group discussion or group consensus condition.

The analyses using individuals as observations yielded no correlations significant at $p = .05$; however, two of the correlations using group rankings were statistically significant and several others approached significance. The rank-order correlations for the group data are shown in Table 3. Composite

TABLE 3

RANK-ORDER CORRELATIONS OF AVERAGE ORIENTATION SCORES WITH NOMINAL GROUP COMPOSITE SCORES, THE REAL GROUP SOLUTION, AND THE DIFFERENCE BETWEEN THE COMPOSITE SCORE AND THE REAL GROUP SCORE

Score	Solution quality scores		
	Nominal group composite	Real group consensus	Difference
Self-orientation	.55*	.45	.37
Interaction orientation	−.68**	−.37	−.46
Task orientation	.27	.07	.25

Note.—$N = 12$.
* $p = .05$.
** $p = .01$.

scores for nominal groups tended to be higher for those groups with high-average self-orientation scores and low interaction orientation scores. However, groups with high-average interaction orientation scores tended to show less of a decrement in problem-solution quality when they were asked to reach a group consensus. Somewhat surprisingly, the task orientation variable yielded the lowest relationships.

No consistent relationships were found with any of the biographical items.

DISCUSSION AND CONCLUSIONS

The results of the present study tend to support the more recent research on individual versus group problem solving cited earlier. Group participation and discussion did not generate higher scores but rather tended to be inhibitory. The quality of the group solution was inferior to the nominal group's composite score and was even inferior to the average individual solution. It should be pointed out that such a finding is dependent on the type of problem used and the manner in which it was scored. Also, nothing has been said about the properties of the group situation which might have led to the reduction in quality scores. Perhaps new and better alternatives were stimulated but were just never expressed.

Groups seemed to settle very early on a particular line of attack, and alternative work procedures were seldom explored in the group discussion. However, this did not seem to occur in an atmosphere of obvious inhibition and threat. Most *S*s knew one another on a first name basis and all seemed highly motivated and enthusiastic. Also, it is doubtful that time limitations could explain the inferiority of the group solution. A 2½-hr. period was set aside for the experimental session and no time limitations were imposed. The absence of such a limitation was reinforced by the experimenter. The experimental sessions were relaxed but task oriented, and none of the *S*s expressed any discomfort because of being away from his job too long. The degree of *S* cooperation and involvement was ideal.

The results obtained with the Orientation Inventory are very tenuous but carry the suggestion that high self- and task orientation and low interaction orientation groups perform better on problems such as this, both when their efforts are combined by an independent party and by consensus. However, high interaction orientation groups seem to suffer less than low interaction orientation groups when asked to reach a consensus.

In sum, the present study casts further doubt on the notion of group facilitation of problem solving and this time with a sample from an appropriate population and with a more realistic experimental problem. However, it should be remembered that the group meeting has a number of other uses besides problem solving. Even when problem solving is the aim, there are criteria other than solution quality on which to judge the effectiveness of the group experience. Maier and Hoffman (1964) have attempted to categorize some common management problems into two classes, those for which finding a solution is the crucial aspect and those for which a number of solutions are possible, but the important consideration is the motivation of the individuals implementing the solution. In the latter case, the effects on solution quality may not be as important as the motivational aspects of group participation.

In addition, the present study and those previously cited have said nothing about the utility of individual versus group solutions. Perhaps it takes considerably more time in an organizational setting to ask the group members to problem solve individually and then evaluate and combine the individual solutions than it does to hold a group meeting. If the cost of this additional time outweighs the benefits of the superior solution, it may not be worth the effort. Conversely, Hoffman (1965) has pointed out that a number of methods exist for enhancing the problem-solving effectiveness of groups. However, the question still remains whether the added effort required to implement such techniques yields solutions which result in the real group being superior to the nominal group. The relative utility of various problem-solving strategies for different types of problems is an area which remains unexplored.

REFERENCES

BASS, B. M. *Research manual: The orientation inventory*. Palo Alto: Consulting Psychologists Press, 1962.

DUNNETTE, M. D., CAMPBELL, J. P., & JAASTAD, K. The effect of group participation on brainstorming effectiveness for two industrial samples. *Journal of Applied Psychology,* 1963, **47**, 30–37.

FAUST, W. L. Group versus individual problem-solving. *Journal of Abnormal and Social Psychology,* 1959, **59**, 68–72.

GERARD, H. G., & MILLER, N. Group dynamics. In P. R. Farnsworth (Ed.), *Annual review of psychology*. Palo Alto: Annual Reviews, 1967.

HALL, E. J., MOUTON, J. S., & BLAKE, R. R. Group problem solving effectiveness under conditions of pooling vs. interaction. *Journal of Social Psychology,* 1963, **59**, 147–157.

HOFFMAN, L. R. Group problem solving. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. (Vol. 2) New York: Academic Press, 1965.

LORGE, I., FOX, D., DAVITZ, J., & BRENNER, M. A survey of studies contrasting the quality of group performance and individual performance. *Psychological Bulletin,* 1958, **53**, 337–372.

MAIER, N. R. F., & HOFFMAN, L. R. Types of problems confronting managers. *Personnel Psychology,* 1964, **17**, 261–269.

MAIER, N. R. F., SOLEM, A. R., & MAIER, A. *Supervisory and executive development*. New York: Wiley, 1957.

TAYLOR, D. W., BERRY, P. C., & BLOCK, C. H. Does group participation when using brainstorming facilitate or inhibit creative thinking? *Administrative Science Quarterly,* 1958, **3**, 23–47.

TUCKMAN, J., & LORGE, I. Individual ability as a function of group superiority. *Human Relations,* 1962, **15**, 45–51.

(Received June 5, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 3, 211–217

ENVIRONMENTAL FRUSTRATION AND CREATIVE PROBLEM SOLVING

BERNARD L. HINTON [1]

Graduate School of Business, Stanford University

Report of an empirical study of the effect of environmental frustration on creative-problem-solving performance. The research design is based upon a linear model of creative problem solving. Alternate forms of analysis are applied, testing the relationship between environmental frustration and creativity under various sets of assumptions about the experimental conditions. It is concluded that environmental frustration does significantly reduce creative-problem-solving performance. This is supported under a number of alternative assumptions about the "rest of the world" variables of personality and ability factors.

Recent research on creative problem solving has tended to focus primarily on the factors of the creative person (e.g., Barron, 1953, 1954, 1955, 1957; Dentler & Mackler, 1964; Drevdahl, 1956; Garwood, 1964; Guilford, 1959; Hitt & Stock, 1965), the creative process (e.g., Crutchfield, 1961; Ghiselin, 1956; Harmon, 1956; Taylor, 1959), or the measurement of creativity through the evaluation of the creative product (e.g., Ghiselin, 1958; McPherson, 1956; Taylor, Smith, & Ghiselin, 1959). The research which has been reported on investigation of the creative environment has been almost exclusively restricted to studies of organizational climate or group climate (e.g., Crutchfield, 1962; Fiedler, 1962; Gibb, 1951; Ziller, Behringer, & Goodchilds, 1962), with considerably less interest being shown in the micro- and psychological climate as it pertains to the individual problem solver.

Maddi (1964), on the basis of his experience, has labeled some generally held assumptions about the effects of psychological stress as being "old wives' tales." One of these assumptions concerns the reduction of creativity in frustrating environments. Maddi argues, in part, that a creative person can ignore, or at least overcome, such restriction and offers case examples to support his position. The research reported herein was designed to test the validity of that position; the null hypothesis is: Creative-problem-solving performance will not be influenced by a state of frustration experienced by the problem solver.

Procedure

A battery of criterion measurement instruments was administered, comprised of an Alternate Uses test, a Match Problems test, and a Consequences test (scored for both "obvious" and "remote" responses). Certain modifications to the published form of these tests were found to be necessary due to the pre-post-test research design. Primarily these required the partitioning of the instruments into two comparable forms. In the case of the Match Problems and the Consequences tests this merely required a physical division of the test booklets. The reliability of the Match Problems test was judged to suffer a negligible amount by such action. The reliability of the Consequences test had already been demonstrated to be unaffected, as the manual reported as few as eight items could be used without affecting the reliability, and after generation of alternate forms there remained 10 items in each form.

In the case of the Alternate Uses test, however, a split into halves was judged to have a serious effect on reliability and thus a parallel set of test questions was deemed necessary. This generated set was balanced, pair-wise, against the original set according to several dimensions, including the nature of the objects listed in the test, the relative size of the objects, and the estimated cost of the objects. A sample pretesting of the response frequency of this set against the original set indicated comparability. To further reduce the introduction of bias or error, the two sets were then merged, with each set having half of its items from the original set and half from the generated set. Additionally, the half from the generated set were the paired items

[1] Now at Indiana University, School of Business. This article is derived, in part, from a doctoral dissertation completed at Stanford University. The author wishes to express appreciation to Charles W. Haley, now at University of Washington, for his assistance on the model development and to Thomas W. Harrell for his many contributions to the research effort.

to those now placed in the alternate form. A random-sample item analysis on response frequency after the original collection of data indicated that the original and generated sets of items were comparable in that no significant differences occurred in the mean number of responses given to the generated items and those given to the original items.

The initial testing was carried out at the beginning of the academic year. The research *Ss* were entering Master of Business Administration (MBA) candidates in the Graduate School of Business at Stanford University. The data were collected prior to registration and in conjunction with the collection of data for other on-going research. The structure of this situation was such as to tap a situation where the level of frustration was probably as low and as randomly distributed as it is possible to achieve in students in the real world. At the least, the environment contained no identifiable sources of frustration. What frustration may have been present, from whatever source, does of course serve to increase the random-error term in the analysis presented later.

Parallel test forms were used for the second administration, which was carried out about 6 wk. later, well into the fall quarter. Prior observation of MBA students had indicated that this point in time was for many of them a peak in their level of frustration as they underwent an adjustment to the high level of output required of them by the academic program. Additionally, the tests were administered immediately after the return of a major paper, a task on which many students had failed to achieve their expected performance level.

Also administered were two sheets of graphic rating scales which were designed to measure the emotional state and attitudes of the testees. Included were distractor questions intended to help mask the prime interest area. The second of these sheets presented eight scale measures on which the *S* indicated what his emotional state had been while taking the tests. The *S*'s response on the frustration scale was of primary interest.

Scoring of the test battery was done after all of the data had been collected, but was independent of *S*'s responses on the graphic scales. (It was "blind" in that the quantification of the scale responses was the last step in the scoring procedure.) This was deemed necessary in that a fair degree of scoring subjectivity was implicit on both the Alternate Uses and Consequences tests. The Match Problems test was scored objectively. The scores were recorded separately on each factor, as a joint score on the Consequences test, and as a total (additive) score. Analysis was carried out on each of these six sets of dependent variable measures.

The administration, in both instances, was carried out in large groups and in a classroom situation. The intent was to hold as constant as possible the physical setting (thus avoiding the introduction of extraneous stimuli or cues) and the sociological setting. The *Ss* were not informed of the nature of the criterion (creativity) until after the testing had been completed, although the nature of the tests may have given some indication of this.

After the collection of the data had been accomplished, as described above, the raw scores on the various creativity measures were transformed into standard scores (*T* scores) with a mean of 50 and a standard deviation of 10. This procedure rendered the scores on the alternate forms of the criterion instruments used directly comparable. The initial sample size was 226, but after the elimination of those *Ss* for whom incomplete data were received, and all of those whose first language was other than English, the final sample size was 147. Data were incomplete on the criterion test battery in only three cases. The remaining *Ss* eliminated because of incomplete data either failed to complete the graphic scales or gave identical responses to each scale.

THE MODEL

This research project was designed through the use of a model of creative problem solving (Hinton, 1966) which has been suggested as offering a basis for the comparison and design of creativity research. The proposed model hypothesized a simple linear relationship between the variables of person, environment, process, and product. Process is considered, however, to be an intervening variable, and is not directly expressed in the model. The model's initial expression is thus:

$$C_{ijk} = \alpha A_{ijk} + \beta P_{ijk} + \gamma E_{ijk} + \theta S^*_{ijk} + \lambda S^{**}_{ijk} + e \quad [1]$$

where C represents an index of creative-problem-solving performance, A an index of given ability or intellect factors, P an index of personality factors, E the psychological or emotional state, S^* the sociological setting, and S^{**} the physical setting. In each case the subscript i identifies an individual, the subscript j the specific variable, and k the point in time.

Expressed verbally, creativity is a function of the person and the environment where the person is represented by measures of his general level of ability and measures of his personality makeup (although these are not necessarily independent), and the environment is comprised of psychological, sociological, and physical elements.

For the current research effort C_{ijk} is defined as the performance or score on the research instruments or tests of creative-problem-solving performance. E_{ijk} is defined as the measured value taken from the eight graphic rating scales on which *S* indicated his emotional state while completing the test battery. There are, therefore, a total of six j's for C_{ijk}, and eight j's for E_{ijk}. Within the time span and structure of this research effort the values of A_{ijk}, P_{ijk}, S^*_{ijk}, and S^{**}_{ijk} are unknown but assumed constant. That is, the change ΔA_{ij}, ΔP_{ij}, ΔS^*_{ij}, and ΔS^{**}_{ij} from the first administration to the second administration is assumed to be essentially zero.

TABLE 1

HYPOTHESIS AND SCHEME OF ANALYSIS

Model statement	Data analysis to be applied*	Assumptions required for unbiased estimate
$C_{ij1} = C_{ij2}$	Simple correlation: C_{ij1}/C_{ij2} Covariance analysis: Variate = C_{ij2}; Covariate = C_{ij1}; Treatment Groups allocated by Emotional State Measure	$A_{ij1} = A_{ij2}$; $P_{ij1} = P_{ij2}$; $C \neq E_{ij2}$
$\Delta C_{ij} = \gamma E_{ij2} + e$	Simple correlation: C_{ij}/E_{ij2} for E_{ij} = (Frustration) and for E_{ij} = (3-scale factor)	$A_{ij1} = A_{ij2}$; $P_{ij1} = P_{ij2}$; E_{ij1} random with respect to E_{ij2}
$C_{ij2} = \delta C_{ij1} + \gamma E_{ij2} + e$	Multiple correlation: $C_{ij2}/C_{ij1}, E_{ij2}$	$A_{ij1} = A_{ij2}$; $P_{ij1} = P_{ij2}$

* Analysis applied is not necessarily direct test of model statement. The analysis tests strength of relationship rather than yielding coefficient for factor in model statement. The results must be interpreted accordingly.

Table 1 shows the model statements which may thus be derived [2] and the analysis to be applied.

It should perhaps be noted that the tests applied were not specifically testing the model statements. That is, the author was interested in the strength or predictive power of the relationship rather than the value of the coefficient. While the model does imply that it is possible to accurately quantify the relationship, it was in fact not realistic to do so, in view of the nature and design of the research and in spite of the fact that such coefficients could readily be calculated.

RESULTS

The emotional state scale measures (quantified values) were analyzed, and the intercorrelations shown in Table 2 determined. Primary interest is upon the frustration measure, but analysis was also carried out on a factor composed of three scales (frustration, discomfort, and dissatisfaction) which show fairly high intercorrelations and predictive power in terms of the criterion measures. Such a factor was presumed to have greater stability than the single measure factor, but the results did not show many significant differences between the two factors.

Following the analysis scheme, as presented in Table 1, correlation coefficients (shown in Table 3) were obtained between the first and second administration scores on the creativity criterion measures.

The correlation between these scores would have an expected value of 1.00 only if the reliability of the measurement instruments were perfect *and* if all other influencing factors were constant in their effect from Time

[2] For the derivation of these statements the reader is referred to Hinton (1966).

TABLE 2

CORRELATION MATRIX—EMOTIONAL STATE SCALE MEASURES

Feeling	Frustration	Discontentment	Discouragement	Inadequacy	Anxiety	Aggression	Discomfort	Dissatisfaction
Frustration	1.000	.172	.660	.552	.434	.054	.562	.625
Discontentment		1.000	.123	.035	.041	.038	.310	.192
Discouragement			1.000	.695	.517	.005	.469	.617
Inadequacy				1.000	.459	.004	.492	.474
Anxiety					1.000	.187	.442	.411
Aggression						1.000	.142	.054
Discomfort							1.000	.530
Dissatisfaction								1.000

TABLE 3

CORRELATION COEFFICIENTS FOR FIRST AND SECOND ADMINISTRATION SCORES ON CREATIVITY CRITERION MEASURES

Criterion measure (C_{ijk})	Correlation coefficient $C_{ijk=1}/C_{ijk=2}$
(j = 1) Alternate uses	.6466
(j = 2) Match problems	.4899
(j = 3) Consequences–obvious	.6104
(j = 4) Consequences–remote	.4133
(j = 5) Consequences–total	.6885
(j = 6) Total score	.7147

1 to Time 2. In view of the reliability estimates (from manual, with range from .70 to .86) and the assumptions concerning variable changes over the time period concerned, the correlations shown do indicate that a change in the environment (the only variable in the model assumed to have changed) had some effect on creativity. In order to test the significance of this, and to relate it directly to frustration, a more direct test of the null hypothesis was carried out using covariance analysis.

The *S*s were divided into three groups according to their response to the scale Frustration on the Emotional State Scale Measure. This assignment was done according to two schemes, the first assigning cases to groups according to an equal-interval subdivision of the frustration-scale responses, the second according to a ranking of the frustration-scale responses, and then a split into three groups of approximately equal size. Ranking ties were not split, thus a minor inequality in group size. These assignments are shown in Table 4.

TABLE 4

CASE ASSIGNMENT FOR COVARIANCE ANALYSIS

F scale value	Group assigned	No. cases
E_{ij2} = 0.00 to 5.00	1	20
E_{ij2} = 5.01 to 9.99	2	75
E_{ij2} = 10.00 to 15.00	3	52
E_{ij2} = 0.00 to 6.20	4	51
E_{ij2} = 6.21 to 9.99	5	44
E_{ij2} = 10.00 to 15.00	6	52

An analysis of covariance was then carried out on these "treatment groups" using the C_{ij2} or creativity criterion test score on the second administration as the variate and the C_{ij1} or creativity test score on the first administration as the covariate. This was done for each of the criterion measures. The results of this analysis are shown in Table 5.

Accordingly, since the analysis shows the differences to be significant, and the only variable presumed to have changed is the emotional state, the null hypothesis is rejected for each of the criterion measures except the Consequences–Total score. That is, the conclusion is drawn that the emotional state, or more specifically the state of frustration, did have an effect on creative-problem-solving performance.

The next test applied evaluates the strength of the relationship between the change in the creativity criterion test score ($\Delta\ C_{ij}$) and E_{ij2}, the emotional state at the time of the second administration, for E_{ij2} = Frustration and for E_{ij2} = Frustration + Discomfort + Dissatisfaction.

Summary statistics for this analysis are shown in Table 6. This table gives both the correlation coefficient and the r^2 (coefficient of determination) for each of the measures. These results are significant at the .05 level or better for all criteria except the Consequences–Obvious.

The next test to be carried out was that of the multiple relationship of C_{ij2} with C_{ij1} and E_{ij2}, that is, the creative-problem-solving score on the second administration as ex-

TABLE 5

COVARIANCE ANALYSIS F RATIOS FOR CREATIVITY CRITERIA

Criterion (C_{ij2})	F ratio for Groups 1, 2, 3	F ratio for Groups 4, 5, 6
(j = 1) Alternate uses	9.099**	9.996**
(j = 2) Match problems	6.193**	5.348**
(j = 3) Consequences–obvious	3.083*	3.683*
(j = 4) Consequences–remote	37.414**	43.049**
(j = 5) Consequences–total	2.986	2.860
(j = 6) Total score	11.247**	11.511**

* $p < .05$; F = 3.07 or more.
** $p < .01$; F = 4.79 or more.

TABLE 6

CORRELATION COEFFICIENTS FOR EMOTIONAL STATE MEASURES AND CHANGE IN CREATIVITY CRITERION MEASURE SCORES

Criterion (ΔC_{ij})	E_{ij2} = Frustration		E_{ij2} = 3-scale variable	
	r	r^2	r	r^2
(j = 1) Alternate uses	−.3019**	.0911	−.3028**	.0917
(j = 2) Match problems	−.2599**	.0676	−.2124*	.0451
(j = 3) Consequences–obvious	.1748*	.0306	.0367	.0013
(j = 4) Consequences–remote	−.5694**	.3242	−.3563**	.1269
(j = 5) Consequences–total	−.1933**	.0374	−.2588**	.0670
(j = 6) Total score	−.3690**	.1362	−.3631**	.1318

* $p < .05$.
** $p < .01$.

plained by the parallel score on the first administration and the effects of the state of frustration indicated by the emotional state measure. This should theoretically give comparable results to that presented in Table 6 except that, whereas the analysis presented in Table 6 assumed the effect of E_{ij1}, the emotional state of the S at the first administration, to be randomly distributed with respect to the emotional state at the second administration (E_{ij2}), no such assumption is required for this analysis. Any effect the emotional state had upon creativity for the first administration is now adjusted or accounted for in the term C_{ij1}. The summary statistics for this analysis are shown in Table 7.

To the degree that the results presented in Tables 6 and 7 differ, such differences may be ascribed to the presence of some degree of frustration at the time of the first administration, some unknown measurement error, and the predictive contribution of C_{ij1}, which has already been seen to be from .41 to .71 (Table 3). For *direct* comparison, the data in Tables 6 and 7 need to be corrected for this. The results in both analyses do indicate that the state of frustration evidenced by S contributes a significant effect to that S's creative-problem-solving performance.

In summary, then, the data analysis has indicated that the emotional state of an individual, and more particularly the state of frustration, does significantly influence his creative-problem-solving performance. This finding is supported by both the covariance analysis presented and by the multiple-regression analysis presented.

DISCUSSION

The objective of this research effort has been an investigation of the effects of an individual's emotional state, specifically his

TABLE 7

MULTIPLE CORRELATION COEFFICIENTS—CREATIVITY CRITERION MEASURE SCORES FOR FIRST AND SECOND ADMINISTRATIONS AND EMOTIONAL STATE MEASURES

Criterion	Criterion (C_{ij1}), E_{ij2} = Frustration		Criterion (C_{ij1}), E_{ij2} = 3-scale variable	
	R	R^2	R	R^2
(j = 1) Alternate uses	.7104	.5047	.6817	.4647
(j = 2) Match problems	.5512	.3038	.5644	.3186
(j = 3) Consequences–obvious	.6303	.3977	.6114	.3738
(j = 4) Consequences–remote	.7017	.4924	.5618	.3156
(j = 5) Consequences–total	.7077	.5008	.7236	.5236
(j = 6) Total score	.7725	.5968	.7707	.5939

state of frustration, on creative-problem-solving performance.

Used as criterion measures have been three tests which tap four separate factors of intellectual ability believed to be basic to creativity. Two additional criterion measures have also been applied, one being the total sum of the scores on the four factors, the other being the sum of two separate scores derived from the Consequences test. Thus a total of six criterion measures have been used. These measures, while only approximations to general creative abilities and not universally accepted, do possess construct validity and have been fairly widely utilized.

Throughout the Results section many of the analyses presented show a very similar pattern of level of significance across these six criterion measures. This pattern is not due to the reliability of the various measures, however, since it does not correspond to the reliability pattern as derived from the test manuals. At least part of the pattern appears to be explained by the relationship between scores. The three measures affected are the Consequences–Obvious, Consequences–Remote, and Consequences–Total.

The Consequences test, it should be remembered, yields two separate factor scores. Also utilized as a criterion score was the sum of these two scores, the Consequences–Total score, reflecting the total number of responses given to the test. As shown in Table 6, the correlation coefficients between the creativity criterion measures and the frustration measure are all negative except for the coefficient for the Consequences—Obvious measure. This indicates that, in general, the presence of frustration resulted in a decrease in creative-problem-solving performance. In the case of the Consequences–Obvious measure, however, the correlation coefficient is seen to be positive. The Consequences–Total score, being the arithmetic sum of the Consequences–Obvious and Consequences–Remote measures, can thus be seen to have suffered from a canceling effect.

That is, performance under a state of frustration relative to performance under the initial and presumed nonfrustrated state is being measured by these nonindependent scores. The total number of responses given to the Consequences test tended to go down under frustration. At the same time, however, the ratio of obvious to remote responses shifted even more drastically, so that the end result was a major decrease in remote responses, a moderate decrease in total responses, but an increase in obvious responses. In a sense the quality of the responses, as well as the quantity, was affected, thus the shift in the sign of the correlation coefficient, and the canceling effect in terms of the tests of significance on the summed variables. It should be noted that this applies to all of the analyses based on these scores, although the magnitude of the effect varies.

The Consequences–Remote criterion measure appears especially strong. This is encouraging in view of the fact that this score measures the factor DMT (Divergent Semantic Transformations), defined as Originality (Guilford & Hoepfner, 1963). This is perhaps the one single factor which most closely approximates the usual, but larger, concept of creativity.

In view of the findings, then, strong support is believed present for the rejection of the null hypothesis, and the conclusion that frustration does reduce problem-solving performance. The correlation between the change in creativity criterion scores under frustration and the emotional state measures evidences r^2's up to .324 in the case of the Consequences–Remote score. While the results in terms of the other criterion measures do not approach this level, it should perhaps be pointed out that this particular analysis is not, by its nature, extremely sensitive since it is not corrected in any way for the possibility of frustration having existed at the time of the first administration, or the influence of the personality variables, or any of the many other factors which may exert an influence.

Analysis of the second administration scores as explained by the first administration scores and the emotional state yielded R^2s (multiple coefficients of determination) ranging from .304 to .597, with the Total score criterion giving the highest coefficient. This indicates some stability to the effect of frustration—that is, it implies that the effect is of similar nature for each of the criterion

measures even though the *size* of the effect may be different.

In hindsight it is obvious that a more optimal procedure than that utilized would have been to collect data on the level of frustration present in *Ss* at this initial point; this was not in fact done. Such collection would have freed one from the necessity of making any assumptions about the level and distribution of the emotional factors, and statistical corrections could have been made to the posttest data to adjust for any variance found.

REFERENCES

BARRON, F. Some personality correlates of independence of judgment. *Journal of Personality,* 1953, **21**, 287–297.

BARRON, F. Some relationships between originality and style of personality. *American Psychologist,* 1954, **9**, 326.

BARRON, F. The disposition toward originality. *Journal of Abnormal and Social Psychology,* 1955, **51**, 478–485.

BARRON, F. Originality in relation to personality and intellect. *Journal of Personality,* 1957, **25**, 730–742.

CRUTCHFIELD, R. The creative process. In, *Conference on the Creative Person.* Berkeley: University of California, Institute of Personality Assessment and Research, 1961. Ch. 6.

CRUTCHFIELD, R. S. Detrimental effects of conformity pressures on creative thinking. *Psychological Beiträge,* 1962, **6**, 463–471.

DENTLER, R. A., & MACKLER, B. Originality: Some social and personal determinants. *Behavioral Science,* 1964, **9**, 1–7.

DREVDAHL, J. E. Factors of importance to creativity. *Journal of Clinical Psychology,* 1956, **12**, 21–26.

FIEDLER, F. E. Leader attitudes, group climate, and group creativity. *Journal of Abnormal and Social Psychology,* 1962, **65**, 308–318.

GARWOOD, D. S. Some personality factors related to creativity in young scientists. *Journal of Abnormal and Social Psychology,* 1964, **68**, 413–419.

GHISELIN, B. The creative process and its relation to the identification of creative talent. In C. Taylor (Ed.), *The 1955 University of Utah Research Conference on the Identification of Creative Scientific Talent.* Salt Lake City: University of Utah Press, 1956. Pp. 195–203.

GHISELIN, B. Ultimate criteria for two levels of creativity. In C. Taylor (Ed.), *The 1957 University of Utah Research Conference on the Identification of Creative Scientific Talent.* Salt Lake City: University of Utah Press, 1958. Pp. 141–155.

GIBB, J. R. The effects of group size and threat reduction on creativity in a problem solving situation. *American Psychologist,* 1951, **6**, 324.

GUILFORD, J. P. Traits of creativity. In H. Anderson (Ed.), *Creativity and its cultivation.* New York: Harper, 1959. Pp. 142–161.

GUILFORD, J. P., & HOEPFNER, R. Current summary of structure-of-intellect factors and suggested tests. Report from the Psychological Laboratory, No. 30, University of Southern California, December 1963.

HARMON, L. R. Social and technological determiners of creativity. In C. Taylor (Ed.), *The 1955 University of Utah Research Conference on the Identification of Creative Scientific Talent.* Salt Lake City: University of Utah Press, 1956. Pp. 42–52.

HINTON, B. L. A model of creative problem solving performance and the effects of frustration. Unpublished doctoral dissertation, Stanford University, 1966.

HITT, W. D., & STOCK, J. R. The relation between psychological characteristics and creative behavior. *The Psychological Record,* 1965, **15**, 133–140.

MADDI, S. R. Motivational aspects of creativity. Chicago: University of Chicago, Department of Psychology, 1964. (Mimeo) (Paper presented at the meeting of the American Psychological Association, Los Angeles, September 1964.)

MCPHERSON, J. H. A proposal for establishing ultimate criteria for measuring creative output. In C. Taylor (Ed.), *The 1955 University of Utah Research Conference on the Identification of Creative Scientific Talent.* Salt Lake City: University of Utah Press, 1956. Pp. 62–68.

TAYLOR, C. W., SMITH, W. R., & GHISELIN, B. Analysis of multiple criteria of creativity and productivity of scientists. In C. Taylor (Ed.), *The 1959 University of Utah Research Conference on the Identification of Creative Scientific Talent.* Salt Lake City: University of Utah Press, 1959. Pp. 5–28.

TAYLOR, I. A. The nature of the creative process. In P. Smith (Ed.), *Creativity.* New York: Hastings House, 1959. Pp. 51–82.

ZILLER, R. C., BEHRINGER, R. D., GOODCHILDS, J. D. Group creativity under conditions of success or failure and variations in group stability. *Journal of Applied Psychology,* 1962, **47**, 34–49.

(Received June 13, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 3, 218–223

LEADERSHIP IN NEGOTIATIONS AND THE COMPLEXITY OF CONCEPTUAL STRUCTURE [1]

SIEGFRIED STREUFERT, SUSAN C. STREUFERT

Purdue University

AND CARL H. CASTORE

Rutgers, The State University

Dyad negotiation teams (homogeneous in the complexity of the individuals' conceptual structure) spent several hours attempting to solve a simulated international conflict situation. Every participant repeatedly rated himself and the other participants on Stogdill's 10 leadership characteristics. The dyad member with the overall highest ratings on all scales was considered the dyad's leader. The effect of simplicity and complexity of conceptual structure on the degree to which leaders display Stogdill's leadership characteristics was analyzed in an ANOVA design. The interaction effect for complexity and leadership characteristics was highly significant. Complex leaders were rated higher on tolerance for uncertainty, assumption of leadership role, consideration, and predictive accuracy. Simple leaders were rated higher on initiation of structure, production emphasis, and demands reconciliation.

In a series of statements, Driver and Streufert (1965, 1966a, 1966b) and Schroder, Driver, and Streufert (1967) have been concerned with complexity theory. According to these authors, individuals differ in the complexity of their conceptual structure, that is, in the degree to which they use differentiation and integration in information processing. The theory suggests that "simple" persons will relate incoming information (input) to decisions (behavioral output) in a 1:1 fashion. They should rarely generate alternative interpretations of the same information (no differentiation) and they should not consider the consequences of the varying interpretations for the potential decisions they might make (no integration). "Complex" persons on the other hand should tend to generate many alternate interpretations of informational input and should relate these alternative interpretations in various ways before reaching a (often tentative) decision (output). Generally, then, complex persons are more open to changes in the environment, adapt more easily, but are less constant and more liable to try pragmatic trial and error behavior (even when this is unnecessary). Usually, complex persons may be described as more "strategic."

Differences in complexity of conceptual structure are not related to differences in intelligence. Correlations of complexity measures and measures of intelligence have generally been near zero.

The theory has generated a considerable amount of research. This research has served to support a number of the theoretical propositions. As predicted by Driver and Streufert (1965, 1966a, 1966b) and Schroder et al. (1967), individual or group differences in the complexity of conceptual structure account for some of the differences in attitude change (Streufert, 1965), perception of others' intentions and strategies (Streufert & Driver, 1965), decision-making characteristics (Streufert & Schroder, 1965), information orientation, information search, and information utilization (Karlins & Lamm, 1967; Stager, 1967; Streufert, Suedfeld, & Driver, 1965; Suedfeld & Streufert, 1966), and other areas concerned with social interaction (Crano & Schroder, 1967; Sieber & Lanzetta, 1964; Tuckman, 1964).

Little interest has been devoted so far to an analysis of the effect of complexity of conceptual structure on leadership characteristics. According to findings of Schroder, Streufert, and Weeden (1964), some relationships between leadership and complexity do exist. In

[1] This research was supported by the Office of Naval Research, Group Psychology Branch.

a preliminary investigation of leadership, these authors found that groups of complex individuals change their leaders more often than groups of simple individuals.

According to the Driver and Streufert and the Schroder, Driver, and Streufert theory, one would not predict that either simple or complex persons would make more effective leaders. However, one would predict that certain forms of leadership would be more likely displayed by complex persons, and other forms of leadership would be more likely displayed by simple persons. For instance, one would predict that complex leaders would be more willing to work with unknown factors, be more empathic, and predict group success better (when the determinants of group success are complex). On the other hand, one would predict that simple leaders would exceed in emphasis on work success, giving direction (rather than empathy), and calling for unity of purpose. Theoretically, one might postulate that the degree to which complexity or simplicity in leaders may be an asset would depend on the particular requirements of the situation with which leader and follower group must cope.

Concerned with the analysis of leadership characteristics, Stogdill (1948, 1957, 1962) proposed 10 characteristics which may be used to describe leaders: (1) tolerance of uncertainty, (2) persuasiveness, (3) initiation of structure, (4) tolerance of freedom of action, (5) role assumption, (6) production emphasis, (7) consideration, (8) predictive accuracy, (9) demands reconciliation, and (10) representation (Stogdill, 1962). These 10 characteristics include some for which differences between complex and simple leaders might be expected (see above). Such differences, however, would occur only if the environment in which group and leader operate is sufficiently complex, so that differences can emerge (cf. Schroder et al., 1967).

The present experiment is concerned with an analysis of the relationship of leadership characteristics as described by Stogdill's 10 leadership scales to the complexity of conceptual structure in leaders, as measured by the sentence completion test (SCT) of Schroder and Streufert (1963) and Schroder et al. (1967).

In its present form, the SCT employs the following stems: "When I am not certain which decision to make," "When someone disagrees with me it usually means," "When I am with two good friends and they express different points of view," "When a situation can be interpreted in several ways." The *Ss* are given 3 min. to complete each stem and write at least three more sentences on the subject. Responses are scored in terms of the multidimensionality and interrelationships of the response. For instance, a response scored as "simple" might be:

> When someone disagrees with me it usually means that he is wrong. I do not take issue when I am not sure of my ground. I try very hard to present the logic of my argument to the other person. If he will not listen I just give up. Some people never learn.

Other simple responses state just as categorically that, "When someone disagrees with me it means that I have made another mistake. . . ."

A complex response to the same stem might be:

> When someone disagrees with me it usually means that we have had different experiences. We see the situation in a different way, and a long discussion is likely to ensue. If we both listen carefully, we will probably find the meaning in the other's point of view, and where it can be applied. Often we both change our views somewhat.

A complex simulated environment, the Tactical and Negotiation Game (TNG; Streufert, Kliger, Castore, & Driver, 1967) was chosen for this research. Leadership is measured in a simulated government (internation) negotiation setting, which, because of the constellation of demands and offers, holds considerable similarity to both governmental and industrial negotiation situations. In addition, this simulation situation provides sufficient environmental complexity to permit individual differences in complexity of conceptual structure to take effect.

METHOD

Subjects

Three hundred and fifty male undergraduate students in an introductory psychology course at an eastern state university were tested for complexity of conceptual structure with the SCT (Schroder et al., 1967; Schroder & Streufert, 1963). The potential

score range of *S*s on this test varies from score 1 (very simple) to score 19 (very complex). The highest score obtained in this sample was 17. The *S*s were selected into simple (score 1 and 2) and complex (score 10 or above) categories. Forty *S*s, 20 of simple conceptual structure and 20 of complex conceptual structure, were selected for this experiment. The *S*s within each of these groups were randomly assigned to dyad negotiation teams, so that two persons on one team would negotiate with two members of the opposing team. Teams were homogeneous in conceptual structure and negotiated with teams of the same level of conceptual structure. Differences in intelligence (College Board scores) between groups were not obtained.

Setting

The environment used for the present experiment was the negotiation component of the TNG described by Streufert et al. (1967). The topic of negotiations was the attempt at a settlement of an international conflict situation (based on the work of Streufert, Clardy, Driver, Karlins, Schroder, & Suedfeld, 1965). The negotiations, in effect, are not unlike potential negotiations for a settlement of a Vietnam or Cuba type conflict.

The *S*s were given 1½ hr. to study the participants' manual for the TNG.[2] The manual contains considerable background information on the negotiations, the demands and offers to be employed by the team members, and potential demands and offers expected from the opposing dyad team (see below). All background information in the manual was phrased in terms intended to persuade *S*s that they were negotiating for the "good" side. The biased presentation of this material was found to be very effective.

The *S*s were presented with 8 offers and 10 demands, both phrased in maximum and minimum terms. They were told that they were expected (by their government) to get the opponent team to accept as many demands as possible, and to make as few offers as possible. The *S*s were informed that their pay for participation in the game would depend directly on the number of offers and demands made. For instance, *S*s would receive $2.00 for each of their maximum demands and $1.50 for each of their minimum demands that was accepted by the enemy team. In turn, they would lose $1.00 for each maximum offer and $.75 for each minimum offer made to and accepted by the enemy team. In addition, they would lose $1.00 to $1.25 (depending on whether maximum or minimum) for each enemy demand they accepted, and gain $1.25 to $1.50 (depending on whether maximum or minimum) for each enemy offer they accepted.

The *S*s were only given general information about the content of enemy demands and offers. They were consequently unable to distinguish between maximum and minimum proposals put forth by the other team.

The demands and offers to be made by the two teams were written so that some of them would be directly contradictory, in other cases the minimum demand of one team would match the maximum offer of the opposing team, and in yet other cases there was no relationship between the demands and offers of the two teams. Without the direct knowledge of the *S*s, demands and offers were organized so that an even trade of demands and offers would yield each person on both teams a take home pay of $8.00. The *S*s were told that they would be able to make anywhere between zero and $20.00. All agreements had to be in writing and signed by all four negotiators.

Measurement and Leadership Characteristics

Ten series of verbal negotiations between sets of two dyad teams (involving four persons at a time) were carried on for three successive periods of 1 hr. each. The *S*s did not know when the final period would be. None of the teams had resolved their differences entirely after the third period, so that an end effect was avoided. After each of the (three) negotiation periods, *S*s were asked to take a recess. The recess period was used to collect a variety of data. Among the measures which *S*s had to complete individually were the 10 scales for leadership characteristics proposed by Stogdill (1962). (See above.) The *S*s were asked to rate their teammate, themselves, and the two members of the opposing dyad team separately on all 10 scales. Seven-point scales were used, marked "much like" and "much unlike" at the scale ends.

Each of the four *S*s in the negotiation sessions was, in addition, rated (at the end of each period) by a trained observer. Mean ratings (across periods and *S*s) for each *S* on each scale were correlated with the mean ratings (across periods) obtained from the trained observer. The obtained correlation of $r = .87$ was taken as an indicator of interrater reliability for the leadership complex. Because of the high correlation, only ratings by *S*s were used for further analysis.

The high correlation between ratings by *S*s and ratings by the observer serves additionally as a check on the meaning of the leadership scales. Scoring instructions defining the meaning of the categories were given as in Table 1.

By specifically scoring for the characteristics as outlined in Table 1, it was attempted to avoid much similarity or reciprocality in scoring categories. The high correlation of *S* ratings with observer ratings was taken as an indicant that *S*s had followed instructions sufficiently well to avoid confounding of the scoring categories.

To select the "leader" of each dyad of *S*s (negotiators), the members of each dyad were com-

[2] The manual has been deposited with the American Documentation Institute. Order Document No. 9244 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $2.25 for microfilm or $5.00 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress. The manual may also be obtained from the senior author.

TABLE 1
LEADERSHIP CHARACTERISTICS AND SCORING INSTRUCTIONS

Leadership characteristic	Scoring instruction
Tolerates uncertainty	The degree to which the subject was willing to work on any topic at hand although there was doubt that the other side would accept the suggestions made.
Persuasiveness	The degree to which the subject attempted and succeeded in getting his view accepted.
Initiation of structure	The degree to which the subject attempted to impose rules or organizational procedures upon the negotiations.
Tolerates freedom of action	The degree to which the subject permitted in himself and others behavior which was extraneous to the task at hand. (Score here based on interpersonal behavior to avoid reciprocality to "initiation of structure.")
Assumes leadership role	The degree to which the subject attempted to guide the discussion in task areas, no matter whether the discussions were held under structured or unstructured conditions.
Production emphasis	The degree to which the subject attempted to get signed agreements, even if these agreements did not touch on areas where the negotiation teams were far apart.
Consideration	The degree to which a subject expressed an understanding of why the other team was demanding (or offering) particular points, even if these points were declared not acceptable by the subject.
Predictive accuracy	The degree to which a subject was able to foretell in advance which agreement would be made, and what terms would be agreed upon.
Demands reconciliation	The degree to which a subject would not tolerate interpersonal (rather than task-based) conflict. (Emphasis here placed on interpersonal side to avoid reciprocality with "tolerates uncertainty.")
Represents group	The degree to which the subject was verbally active and argued primarily for the demands and offers of his side. (Emphasis on task orientation to avoid similarity to "assumes leadership role.")

pared on a leadership score obtained by calculating the mean of four leadership ratings (by themselves and the three other negotiators) across scales and periods. The person with the highest score on each team was designated the leader. In other words, a leader in this sense is the dyad member with the highest average rating across all of Stogdill's leadership characteristics over an extended period of time.

Since one member of each dyad is designated leader, the 40 *S*s yielded 20 leaders. Ten of these leaders were of complex conceptual structure, the other 10 were of simple conceptual structure. Mean leadership scores for each one of Stogdill's 10 scales (across periods and across ratings by all members of both negotiating dyads) were calculated to use as entries in a mixed-design two-way analysis of variance.

RESULTS AND DISCUSSION

The data were analyzed with a mixed two-way analysis of variance. The main effect for complexity was not significant. The main effect for leadership scales ($F = 62.01$) and the interaction effect of leadership scales with complexity of conceptual structure ($F = 37.74$) were significant beyond the .001 level.

In view of the highly significant interaction, differences between simple and complex leaders on each of the 10 leadership scales were tested by means of the F statistic with 1/168 degrees of freedom. The analysis-of-variance within-error term was used as the estimate of mean square error within (cf. Winer, 1962). Because of repeated application, and because of the small error mean square (below the "theoretical minimum" of 1.0), a limit of $p < .001$ was set to accept significance. Significant F ratios were obtained for tolerance of uncertainty ($F = 128.0$), initiation of structure ($F = 34.03$), assumption of leadership role ($F = 18.0$), production emphasis ($F = 63.28$), consideration ($F = 52.53$), predictive accuracy ($F = 69.03$), and demands reconciliation ($F = 24.75$). The F ratios for persuasiveness, tolerance for freedom of action, and represents group were not significant. The results are shown in graphic form in Figure 1.

As expected, the main effect for complexity was not significant. It appears that leadership characteristics are displayed by both simple and complex leaders. Although it is difficult to interpret the main effect for leadership scales since the interaction effect is highly significant, one might suggest that this

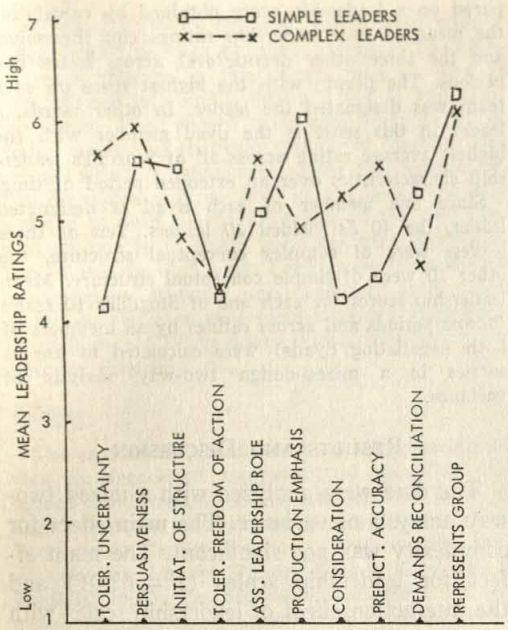


FIG. 1. Mean leadership ratings on Stogdill's characteristics for leaders of simple and complex conceptual structure.

result may be an effect of the negotiation situation. For instance, secrecy and united action are of vital importance in negotiations, and great freedom of action for the follower cannot be permitted by the leader if negotiations are to be successful. Consequently, one finds that both complex and simple leaders are rated near the scale midpoint on this characteristic. Similarly, both are rated high in assuming the leadership role and representing their group, in other words, permitting little (divergent) activity by their follower.

The highly significant interaction effect is of special interest. As expected, there are some communalities and some divergencies in the characteristics of simple and complex leaders. Complex leaders are high on tolerance of uncertainty, while simple leaders are rated near the scale midpoint. Since complex persons have more facilities for integrating incongruent or unstructured information, such a result may be expected. Similarly, it is not surprising that simple leaders are more concerned with initiation of structure. Stogdill describes this characteristic as "clarifies own role and lets follower group know what is expected." This again would result in less uncertainty, establish clear-cut roles, and permit greater emphasis on accomplishing the task as defined. In line with this argument, simple leaders also scored higher than complex leaders on production emphasis. On the other hand, complex leaders show more consideration and demand less reconciliation than simple leaders. Again, this would be due to their greater ability to integrate alternate views, possibly even a preference for alternate views that might serve to develop alternate or more complex strategies. In addition, complex leaders were rated as having higher predictive accuracy. This result corroborates the finding of Streufert and Driver (1965) that complexity is of great advantage in the perception of others' strategy. The finding that complex *S*s are rated higher in assumption of leadership role is difficult to interpret since the terminology is, in this case, somewhat unclear. Further research on the meaning and implications of the leadership "role" seems advisable.

One should note that there was no difference ($r = .85, .89$) in the correlation of the trained observer's ratings with the mean group ratings of complex and simple leaders. In other words, the differences between complex and simple leaders on these scales are not likely due to different uses of the rating scales by complex and simple persons, but are reflections of actual differences in leader characteristics.

To check whether the ratings obtained by the "leader" *S*s were based on actual leadership (rather than other characteristics, e.g., possibly those inherent in the situation), mean ratings were also calculated for "follower" *S*s. Mean ratings on all categories were below the scale midpoint (4.0), and the peaks and valleys of the curves were less pronounced. On the basis of this information it was concluded that validity of leadership ratings was given.

The results suggest that, although in some situations leaders who are complex and leaders who are simple will do equally well, there are other situations requiring specific leadership characteristics in which the simple leaders may outperform complex leaders and vice versa. It may be of value to study the type

of leadership required before a particular leader is selected. In this fashion, leadership and consequently group performance may be optimized. In negotiations, for instance, it may be of value to use a complex leader when the situation is somewhat unstructured and frequent modification of strategies as well as empathy with the opposing team are required. On the other hand, a simple leader might be more advantageous if more unilateral negotiations with emphasis on unmodified results are required. However, a warning should be added: Negotiations between teams whose members are of simple conceptual structure may well end in a breakdown of the negotiations as long as both sides are of similar strength and if there is little or no meeting ground. Under such conditions, one would expect better results by using more complex leaders and negotiations teams.

REFERENCES

CRANO, W. D., & SCHRODER, H. M. Complexity of attitude structure and processes of conflict reduction. *Journal of Personality and Social Psychology,* 1967, **5**, 110–114.

DRIVER, M. J., & STREUFERT, S. The general incongruity adaptation level (GIAL) hypothesis: An analysis and integration of cognitive approaches to motivation. Publication No. 114. Lafayette, Ind.: Institute for Research in the Behavioral, Economic, and Management Sciences, Purdue University, 1965.

DRIVER, M. J., & STREUFERT, S. The general incongruity adaption level (GIAL) hypothesis-II. Incongruity motivation in relation to affect, cognition, and activation-arousal theory. Institute paper No. 148. Lafayette, Ind.: Institute for Research in the Behavioral, Economic, and Management Sciences, Purdue University, 1966. (a)

DRIVER, M. J., & STREUFERT, S. Group composition, input load and group information processing. Institute paper No. 142. Lafayette, Ind.: Institute for Research in the Behavioral, Economic, and Management Sciences, Herman C. Kraennert Graduate School of Industrial Administration, Purdue University, 1966. (b)

KARLINS, M., & LAMM, H. Information search as a function of conceptual structure in a complex problem-solving task. *Journal of Personality and Social Psychology,* 1967, **5**, 456–459.

SCHRODER, H. M., DRIVER, M. J., & STREUFERT, S. *Human information processing.* New York: Holt, Rinehart & Winston, 1967.

SCHRODER, H. M., & STREUFERT, S. The measurement of four systems of personality structure varying in level of abstractness: Sentence completion method. Princeton University: Office of Naval Research Tech. Rep. No. 11, 1963.

SCHRODER, H. M., STREUFERT, S., & WEEDEN, D. C. The effect of structural abstractness in interpersonal stimuli on the leadership role. Princeton University: Office of Naval Research Tech. Rep. No. 3, 1964.

SIEBER, J. E., & LANZETTA, J. T. Conflict and conceptual structure as determinants of decision making behavior. *Journal of Personality,* 1964, **32**, 622–641.

STAGER, P. Conceptual level as a composition variable in small group decision-making. *Journal of Personality and Social Psychology,* 1967, **5**, 152–161.

STOGDILL, R. M. Methods for determining patterns of leadership behavior in relation to organization structure and objectives. *Journal of Applied Psychology,* 1948, **32**, 286–291.

STOGDILL, R. M., & COONS, A. E. (Eds.) Leader behavior: Its description and measurement. Columbus: Ohio State University, Bureau of Business Research, Monograph No. 88, 1957.

STOGDILL, R. M. New leader behavior description subscales. *Journal of Psychology,* 1962, **54**, 259–269.

STREUFERT, S. Communicator importance and interpersonal attitudes toward conforming and deviant group members. *Journal of Personality and Social Psychology,* 1965, **2**, 242–246.

STREUFERT, S., CLARDY, M. A., DRIVER, M. J., KARLINS, M., SCHRODER, H. M., & SUEDFELD, P. A tactical game for the analysis of complex decision making in individuals and groups. *Psychological Reports,* 1965, **17**, 723–729.

STREUFERT, S., & DRIVER, M. J. Conceptual structure, information load and perceptual complexity. *Psychonomic Science,* 1965, **3**, 249–250.

STREUFERT, S., KLIGER, S. C., CASTORE, C. H., & DRIVER, M. J. A tactical and negotiations game for the analysis of decision integration across decision areas. *Psychological Reports,* 1967, **20**, 155–157.

STREUFERT, S., & SCHRODER, H. M. Conceptual structure, environmental complexity and task performance. *Journal of Experimental Research in Personality,* 1965, **1**, 132–137.

STREUFERT, S., SUEDFELD, P., & DRIVER, M. J. Conceptual structure, information search and information utilization. *Journal of Personality and Social Psychology,* 1965, **2**, 736–740.

SUEDFELD, P., & STREUFERT, S. Information search as a function of conceptual and environmental complexity. *Psychonomic Science,* 1966, **4**, 351–352.

TUCKMAN, B. W. Personality structure, group composition, and group functioning. *Sociometry,* 1964, **27**, 469–487.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received June 16, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 3, 224–226

REACTION TIME AS A FUNCTION OF THE CUE PROPERTIES OF AN AUDITORY DISPLAY

J. RICHARD SIMON

University of Iowa

An attempt was made to increase the speed of processing verbal directional commands by using the ear stimulated to provide a relevant directional cue. In a choice reaction-time task, Ss pressed the right- or left-hand key in response to: (1) binaural "right" or "left" commands, (2) monaural "right" or "left" commands in the ear corresponding to the content of the command, and (3) monaural pure tones. Responses to the binaural commands were significantly slower than to the other 2 conditions, and responses to the monaural commands were significantly slower than to the monaural tones.

A basic and continuing problem in engineering psychology is that of designing displays which effectively communicate spatial relationships such as direction (Fitts, 1951). In order to maximize the speed of information processing, displays should be designed so that they provide cues for the direct perception of relevant directional relationships. This occurs, for example, when the spatial arrangement of a visual display is made to correspond with that of its associated control (Fitts & Seeger, 1953; Garvey & Mitnick, 1955).

Simon and Rudell (1967) demonstrated that directional cues also play a role in the interpretation of auditory displays. Their *S*s performed a choice reaction-time (RT) task which involved pressing the right- or left-hand key in response to the commands "right" or "left" which were presented to the right or left ear. The *S*s responded significantly faster when the content of the command corresponded to the ear stimulated (i.e., "right" in right or "left" in left ear) than when it did not (i.e., "right" in left ear or "left" in right ear). In other words, a delay in responding occurred when the irrelevant directional cue (ear stimulated) conflicted with the relevant symbolic cue (content of the command).

The present study represents an attempt to increase the speed of processing verbal directional commands by utilizing the directional cue provided by the ear stimulated. Will *S*s respond faster to a display which combines directional and symbolic cues, that is, when the "right" command is always presented to the right ear and the "left" command to the left ear, than they will to a display which utilizes either directional or symbolic cues alone? The expectation that such a result might occur is based upon evidence that two simultaneous stimuli may summate to elicit a faster response than any one of the stimuli alone (Woodworth & Schlosberg, 1954). Todd (1912), for example, has shown that RT to a combination of sound and shock is faster than to either one alone.

METHOD

Apparatus. The apparatus provided a measure of choice RT to a series of auditory stimuli which were presented to *S* through Grason-Stadler TDH-39 matched earphones. Two telegraph keys were positioned 10 in. apart on a table in front of *S*. He was instructed to depress the correct one of the two keys as quickly as possible after hearing the signal. The instructions and stimulus materials were recorded on tape and presented at approximately 80 db. by a Sony TC 500 Stereocorder. Onset of the signal activated a Hunter Klockounter which stopped when *S* pressed the key. Depressing the key also signaled *E* as to which key (or keys) had been pressed.

Subjects. The *S*s were 18 male and 18 female University of Iowa undergraduate volunteers between the ages of 18 and 29 (average age 19.8). All *S*s passed a standard audiometric screening test in which tones of 500, 1000, and 2000 cps were presented to each ear separately. No *S* had a hearing loss greater than 10 db. at any one of the three frequencies tested.

Procedure and experimental design. Each *S* performed on three blocks of trials corresponding to the three experimental conditions. A 3 × 3 Latin square was used to balance the order of presentation of the conditions. Eight males and eight females

were randomly assigned to each of the three sequences.

In one block of trials, the stimulus was a 1000-cps monaural tone. The *S* was told that he would hear the tone in either his right or left ear and that he was to respond by depressing the corresponding (right or left) key. In a second block, *S* was told that he would hear either the word "right" or the word "left" in *both* ears simultaneously and that he should respond by depressing the appropriate key. In a third block, *S* was told that he would hear the word "right" in his right ear or the word "left" in his left ear and that he was to respond by depressing the appropriate key. Each block of trials consisted of a predetermined random sequence of 42 tones or commands half of which called for a right response and half a left response. The first four responses in each block, two right and two left responses, were practice trials. A ready signal was presented 2 sec. prior to each stimulus and consisted of the simultaneous onset of a warning light and a 2000-cps, 80-db. binaural tone of 500-msec. duration. There was a 7-sec. interval between commands.

RESULTS

Median RTs were computed for each *S* for each treatment condition. The means of these median RTs were as follows: monaural tone (directional cue) = 307 msec.; monaural verbal command (directional plus symbolic cues) = 322 msec.; and binaural verbal command (symbolic cue) = 420 msec. An overall analysis of variance revealed that the difference between these treatment conditions was significant; $F(2/60) = 155.10$, $p < .001$. None of the other main effects (i.e., order, sequence, sex) or interactions was significant.

A critical difference test ($p = .05$) between individual pairs of means revealed that RT to the binaural verbal command was significantly slower than to either of the other two treatment conditions. The RT to the monaural verbal command was significantly slower than to the monaural tone.

DISCUSSION

Results indicated that *S*s were slower in responding to the binaural verbal command than to either the monaural tone or the monaural verbal command. The explanation for this finding seems obvious. With the binaural command, *S* could not begin to respond until an intelligible portion of the word had been articulated, whereas, with the other two treatment conditions, the ear stimulated provided a relevant cue and *S* could begin to respond immediately upon onset of the stimulus. It is therefore not surprising that the monaural verbal command (combined directional and symbolic cues) resulted in a faster RT than the binaural verbal command (symbolic cue). What is surprising is that the monaural verbal command actually resulted in a significantly slower RT than the monaural tone (directional cue). Results indicated clearly that *S*s were utilizing the directional cue contained in the monaural verbal command. Apparently, however, they were not utilizing the cue as effectively as they might have. It was as if the availability of the redundant symbolic cue (content of command) slowed down *S*s' reactions to the directional cue (ear stimulated).

Results of the present study show that RT to verbal commands of "right" or "left" may be reduced by presenting these commands to the appropriate ear, that is, adding a relevant directional cue. However, results also indicate that this method of conveying directional information is not superior and, in fact, is inferior to presenting pure tones to one ear or the other. These results must not be used as a basis for generalizations concerning (*a*) the relative effectiveness of directional versus symbolic cues in auditory RT, or (*b*) the general effect of combining directional and symbolic cues. Such generalizations are unwarranted because of the qualitative difference between the two kinds of stimuli (tone and verbal command) used in this study. Future research should hold this variable constant, for example, train *S*s to associate different directional responses to tones of differing frequency and then compare RT to binaural and monaural presentations of these tones.

REFERENCES

FITTS, P. M. Engineering psychology and equipment design. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1287–1340.

FITTS, P. M., & SEEGER, C. M. S-R compatibility: Spatial characteristics of stimulus and response

codes. *Journal of Experimental Psychology,* 1953, **46,** 199–210.

GARVEY, W. D., & MITNICK, L. L. Effect of additional spatial references on display-control relationships. *Journal of Experimental Psychology,* 1955, **50,** 276–282.

SIMON, J. R., & RUDELL, A. P. Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology,* 1967, **51,** 300–304.

TODD, J. W. Reaction to multiple stimuli. *Archives of Psychology,* 1912, No. 25, 1–65.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1954.

(Received June 29, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 3, 227–233

COMPARISON OF JUDGMENTAL METHODS WITH ENDORSEMENTS IN THE ASSESSMENT OF PERSONALITY TRAITS [1]

IGOR KUSYSZYN

York University, Downsview, Ontario

The study focuses on the degree to which judgment methods are free from dissimulation effects as compared with endorsements. 8 structured personality scales from the Personality Research Form were administered under 4 instructional sets—Endorsement, Desirable in Others, Frequency of Endorsement, and Desirable in Self—to 94 members of 7 local college fraternities. The validity of both judgments and endorsements was examined by relating them to noninventory behavioral criteria in the form of behavior ratings. Although judgments were found to be relatively free from the kinds of dissimulation effects found with endorsements, judgments yielded significantly lower validities than endorsements.

It has long been known that personality tests of the paper-and-pencil variety are subject to faking. Although some tests have employed validity scales to detect distortion, an invalid profile is of little use to the investigator who is interested in the individual's personality traits. One approach to the problem of faking, implemented in the present study, is to change the basic task of the *S* from that of describing himself to that of making judgments about the test items. For example, instead of asking *S* to evaluate an item as "true" or "false" of himself, one might ask him to judge how desirable he thinks others feel it is to respond "true" to the item. Thus, through the use of judgments, the purpose of the task may be disguised so as to minimize the operation of dissimulation tendencies. At the same time judgments may yield enough systematic response variance to provide an indication of the individual's own personality (cf. Campbell, 1950).

There have been very few studies published in which an attempt was made to use judgments for the purpose of assessing personality traits. In an early study, Rosen (1956) administered the Minnesota Multiphasic Personality Inventory under two judgment sets, "personal desirability" and "perceived social desirability," and obtained positive but low correlations between these sets and the standard true-false set. Rosen concluded that the correlations were sufficiently low as to suggest the relative independence of each set. More recently, Messick (1960) factor analyzed social desirability ratings of the Edwards Personal Preference Schedule (Edwards, 1954) and obtained nine dimensions of social desirability. Jackson (1961, 1964a) suggested that the presence of these different "points of view" about desirability raises the possibility that the judgments of items might be useful in drawing inferences about the personality of the judge. Jackson (1961, 1964a) tested this hypothesis by using desirability judgments of a conformity scale to predict an attitude-shift criterion of conformity. The judgments were found to have a slightly higher relationship with the criterion (.29) than the endorsements (.22). In a replication of the Jackson study by Goldberg and Rorer (1966), in which *S*s predicted the responses of their group members instead of judging the desirability of the

[1] This study is based on a thesis submitted by the author in 1967 to the Department of Psychology, University of Western Ontario, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The study has profited greatly from the author's many discussions with Douglas N. Jackson, as well as from his careful reading of the manuscript. The able assistance of J. Edwin Boyd in programming the data for computer analysis is also gratefully acknowledged. This research was supported in part by the University of Western Ontario Research Fund and in part by Grant No. 151 from the Ontario Mental Health Foundation to Douglas N. Jackson.

items, a correlation of .30 between predictions and the criterion was obtained while the validity of the endorsements was only .06. However, in an earlier replication of the Jackson study by Loomis and Spilka (1963), the validity of endorsements (.26) was superior to the validity of the judgments (.15). In a broader study of conformity, Stricker, Messick, and Jackson (1966) related judgments and endorsements of an Independence-Yielding scale to four kinds of conformity criteria and found judgments to be better predictors than endorsements in the case of *S*s who were suspicious as to the nature of the study.

The positive results in favor of judgments in three of the last four studies cited above suggest the potential importance of judgment instructions in trait assessment, and the limited scope of these studies (i.e., only one kind of criterion—conformity—was predicted) clearly indicates the need for a broader and more thorough investigation of judgment methods. The present study will attempt to partially fulfill this need by using criteria which differ in method from questionnaires. These are in the form of behavior ratings by peers. The validities of eight personality scales administered under three kinds of judgment sets and under an endorsement set will be evaluated. The susceptibility of judgments and endorsements to dissimulation effects will be studied by comparing the validities of defensive and social desirability respondents with the validities obtained by low defensive and low social desirability respondents, respectively.

METHOD

Subjects

The *S*s were 94 males ranging from freshmen to graduate students from seven college fraternities in the city of London, Ontario. Thirty-six lived inside their fraternity houses (live-ins) and 58 lived outside (live-outs). They had been acquainted with one another for a mean period of 18 mo. and had a mean age of 21.9 yr., with a standard deviation of 1.7 yr.

Personality Inventory

The inventory consisted of eight 20-item personality scales from the Personality Research Form (PRF) (Jackson, 1964b) and two 20-item validity scales, Social Desirability (Jackson, 1964b) and Defensiveness (Jackson & Messick, 1964). The personality scales are as follows: Nurturance, Exhibition, Autonomy, Play, Aggression, Achievement, Impulsivity, and Order. The scales have a median test-retest reliability of .82 and a median validity, based on behavior ratings, of .54. The social desirability factor was suppressed in all scales during item analysis. Acquiescence is suppressed by balanced true-false item keying. The Social Desirability scale measures the degree to which an individual responds in a socially desirable manner; the Defensiveness scale measures the degree to which a bias in self-report is operating.

Trait Adjectives

For the behavior rating task, eight matched pairs of trait adjectives were selected so as to represent the eight personality scales of the inventory (matched pairs instead of single adjectives were selected so that the reliability of the ratings could be determined). A brief definition of the trait accompanied each adjective. The adjectives and the scales they represent are as follows: supporting, helpful (Nurturance); attention seeking, entertaining (Exhibition); independent, nonconforming (Autonomy); playful, fun loving (Play); aggressive, argumentative (Aggression); achieving, ambitious (Achievement); impulsive, impatient (Impulsivity); orderly, methodical (Order).

The rating instructions asked *S*s to rate how characteristic each trait was of each of the members of their fraternity group. The ratings were made on a 9-point scale ranging from "extremely characteristic" through "neutral" to "extremely uncharacteristic" of the individual.

Instructional Sets

The personality inventory was administered under four instructional sets—one endorsement set and three judgment sets. The sets, briefly defined, are as follows: Endorsement—"Rate how characteristic each statement is of you"; Desirable in Others—"Rate how desirable each statement is assuming it is true of others"; Frequency of Endorsement—"Rate how frequently people in general would answer 'true' to the statement"; Desirable in Self—"Rate how desirable the statement is when it is applied to you." All ratings were made on a 9-point scale similar to that used with the trait adjectives.

Administrative Procedure

Each fraternity group was run separately in a classroom setting. The groups ranged in size from 9 to 19 *S*s. The *S*s performed the adjective trait rating task first, then the personality inventory was administered successively under each of the four instructional sets in the order in which they are described above.

RESULTS

Defensiveness and Social Desirability Groups

To investigate the effects of biased responding, the total sample was split at the

median of the *Ss*' Defensiveness scale scores obtained under the Endorsement set. This yielded a mean Defensiveness scale score of 99 (the possible range of scores being 20–180) for the High Defensive group ($N = 47$) and a score of 76 for the Low Defensive group ($N = 47$). To investigate the effects of socially desirable responding, the total sample was similarly split at the median of the Social Desirability scale scores. This split yielded a mean Social Desirability scale score of 140 for the High Social Desirability group ($N = 47$) and a mean score of 118 for the Low Social Desirability group ($N = 47$).

Reliability of Behavior Ratings

The median matched-pair reliability of the eight pairs of trait adjectives was .88 for the total sample ($N = 94$), .83 for the live-ins ($N = 36$), .79 for the live-outs ($N = 58$), .88 and .86 for the Low Defensive ($N = 47$) and High Defensive ($N = 47$) groups, respectively, and .88 and .88 for the Low Social Desirability ($N = 47$) and High Social Desirability ($N = 47$) groups, respectively (coefficients are Spearman-Brown corrected).

The median between-rater reliability (agreement between subgroups of raters rather than the correspondence between matched pairs of adjectives) for the 16 single adjectives was .83 (Spearman-Brown corrected) for the total sample ($N = 94$).

Reliability of Inventory Scales

The reliabilities of the judgment and endorsement sets did not differ significantly. For all seven groups and for all four instructional sets, the median reliabilities of the 10 inventory scales ranged between .67 and .86 (split-half and Spearman-Brown corrected).

Comparison of Judgment and Endorsement Validity Coefficients

The criterion behavior rating scores were calculated as follows. Each *S* was rated by his peers on each of the 16 trait adjectives. The mean rated score for each *S* for each of the 16 adjectives was then determined. Since the 16 adjectives represented eight matched pairs of adjectives, the means of each of the eight matched pairs were added, yielding eight criterion scores per *S*, each of the eight scores corresponding to one of the eight personality scales of the inventory. Validity coefficients were determined for each of the eight personality scales under each of the four instructional sets by correlating the *Ss*' inventory scale scores with their behavior rating scores. For purposes of taking into account the varying reliabilities of the trait ratings and of the inventory scales, all validity coefficients were initially corrected for attenuation. However, since the corrected coefficients revealed the same patterns of relationships among validities as did the raw coefficients, only the raw validity coefficients will be presented.

Table 1 presents the validity coefficients for the total sample ($N = 94$). The Endorsement set yields the highest validities ($Mdn\ r = .40$), the Desirable in Others set, the next highest ($Mdn\ r = .20$), the Desirable in Self set, the next highest ($Mdn\ r = .17$), and the Frequency set, the lowest ($Mdn\ r = .06$).

For purposes of conserving space, only the median validity coefficients will be presented

TABLE 1

VALIDITY COEFFICIENTS FOR THE TOTAL SAMPLE

	Nurt	Exh	Auto	Play	Agg	Ach	Imp	Order	*Mdn*
Endorsements	29	56	24	58	35	44	32	51	40
Desirable in Others	19	26	10	23	36	01	21	12	20
Frequency	19	−07	−06	02	19	09	−02	09	06
Desirable in Self	19	31	05	37	21	12	01	14	17

Note.—$N = 94$. Decimals omitted.
The .05 and .01 significance levels of r are .21 and .27, respectively.
Significant differences between the Endorsement and Judgment sets as determined by the Wilcoxon signed-rank test: Endorsements—Desirable in Others, $p < .02$; Endorsements—Frequency, $p < .01$; Endorsements—Desirable in Self, $p < .01$.
Abbreviations: Nurt = Nurturance; Exh = Exhibition; Auto = Autonomy; Play = Play; Agg = Aggression; Ach = Achievement; Imp = Impulsivity; Order = Order.

TABLE 2

MEDIAN VALIDITY COEFFICIENTS OF EIGHT PERSONALITY SCALES UNDER JUDGMENT AND ENDORSEMENT SETS

Instructional set	Live-ins[a]	Live-outs[b]	High defensive[c]	Low defensive[c]	High desirability[c]	Low desirability[c]
Endorsement	47	34	37	32	36	48
Desirable in Others	11	17	13	24	18	26
Frequency	14	−03	−01	18	07	13
Desirable in Self	17	10	07	28	07	18

Note.—Decimals omitted.
[a] $N = 36$.
[b] $N = 58$.
[c] $N = 47$.

for the other six groups which were studied.[2] These correlations are presented in Table 2. From Table 2 it is seen that in each group the Endorsement set yields higher validities than any one of the three judgment sets. The Wilcoxon matched-pairs signed-rank test revealed that each of the three judgment sets differed significantly from the Endorsement set in all six groups.

When the endorsement validities of each pair of subgroups were compared (Mann-Whitney U test) live-ins obtained significantly higher validities than live-outs ($p < .05$); the validities of the Defensiveness subgroups and the validities of the Social Desirability subgroups did not differ significantly. When the validities of each pair of subgroups were compared under each of the three judgment sets (Mann-Whitney U test) no significant differences were found between the live-ins and the live-outs and no significant differences were found between the Social Desirability subgroups. However, the Low Defensive group obtained significantly higher validities than the High Defensive group under the Desirable in Others ($p < .01$) and under the Frequency ($p < .01$) sets.

When the individual validity coefficients of Table 1 are examined they are seen to vary greatly in magnitude, especially under the judgment conditions. This variability indicates that certain Judgment Set × Scale Content interactive effects must be operating. That is, the same item content is rated differently under different judgment sets. When the individual validity coefficients of the Defensiveness subgroups were examined, certain Scale × Subgroup interactive effects were also seen to be operating, especially under the judgment sets. For example, for the traits Exhibition, Aggression, and Achievement, Low Defensive *S*s obtained positive validity coefficients while High Defensive *S*s obtained negative validity coefficients, under the same judgment sets. Similar interactive effects were found for the Social Desirability subgroups.

Behavior Rating and Personality Scale Score Means

Table 3 presents the mean behavior rating scores and the mean personality scale scores of the High and Low Defensive groups. These means are compared in the hope that they will shed some light on the pattern of validity coefficients obtained by these two groups. (It will be remembered that the Low Defensive group obtained significantly higher validity coefficients than the High Defensive group on two of the judgment sets—see above.) When the mean scores are examined on the basis of a general "favorable-unfavorable" trait dimension, it is seen that High Defensive *S*s obtain higher scores than Low Defensive *S*s on favorable traits and lower scores

[2] The entire 48 × 48 correlation matrix including all validity coefficients and all off-diagonal correlations has been deposited with the American Documentation Institute. Order Document No. 9828 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.25 for photocopies or $1.25 for microfilm and make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 3

MEAN TRAIT RATING AND MEAN PERSONALITY SCALE SCORES FOR HIGH AND LOW DEFENSIVE GROUPS

	Nurt	Exh	Auto	Play	Agg	Ach	Imp	Order
High defensive group (N = 47)								
Trait ratings	12.34	9.93**	10.14	11.19**	8.05	13.62**	9.86	12.71**
Endorsements	119*	98**	95	104**	85**	116**	93**	116**
Desirability in Others	130	100**	95	98	71	131	81	127
Frequency	114	99	98	105	83	108	94	109
Desirability in Self	132	100**	97	94*	63	137	76	131
Low defensive group (N = 47)								
Trait ratings	11.96	10.97**	10.64	12.64**	8.45	12.65**	10.41	11.55**
Endorsements	113*	114**	94	115**	97**	106**	107**	103**
Desirability in Others	130	107**	94	102	71	131	82	128
Frequency	116	99	95	105	86	111	91	116
Desirability in Self	129	110**	99	100*	68	137	80	129

Note.—The possible score range for a trait rating score is 2–18 and the range for a personality scale score is 20–180. See Note of Table 1 for abbreviations.
The entries marked with asterisks indicate statistically significant differences between the High and Low Defensive groups between the scales obtained under parallel instructional sets.
* $p < .05$.
** $p < .01$.

on unfavorable traits. For example, when the behavior rating scores of the two groups are compared it is seen that the High group obtains significantly lower scores than the Low group on Exhibition and Play (Play was defined as "horsing around," always having a good time, etc.) and higher scores than the Low group on Achievement and Order. Under the Endorsement condition the High group again obtains lower scores on Exhibition and Play and higher scores on Achievement and Order, as well as obtaining lower scores on Aggression and Impulsivity and a higher score on Nurturance. However, High Defensive *S*s do not respond much more favorably than Low Defensive *S*s under the judgment sets: there were no significant differences between the two groups on the Frequency set, only one significant difference under the Desirable in Others set (Exhibition), and only two significant differences under the Desirable in Self set (Exhibition and Play) (see Table 3).

Similar score differences were obtained between the means of the Social Desirability subgroups but fewer of the differences were found to be statistically significant.

DISCUSSION

It will be recalled that the major purpose of the study was to examine the validities of judgments and endorsements, especially in relation to the operation of dissimulation effects. The obtained results were in one sense supporting of the use of judgment sets for controlling response biases, but it became clear that much more work is needed to establish an unequivocal foundation for the use of judgments routinely in personality assessment. Thus, the judgment sets were found to be relatively free from dissimulation effects in that little evidence of biased judgments was found under any of the three judgment sets. However, the validities of the judgment sets were low in relation to the validities obtained with endorsements. Endorsements yielded higher validities for all groups, even in the case of High Defensive and High Social Desirability respondents who endorsed items in a favorable direction.

When the mean personality scale scores of the High and Low Defensive groups were compared under the Endorsement condition,

High Defensive *Ss* obtained more favorable scores than Low Defensive *Ss*. However, it was also observed that High Defensive *Ss* obtained more favorable behavior ratings than Low Defensive *Ss*. That is, not only do High Defensive individuals respond in a favorable direction to personality items, but they also seem to create, quite strongly, favorable impressions upon others in real life. This finding is important for several reasons. Although other factors may be operating, the fact that there is a correspondence between the favorable endorsement of personality items and favorable responding in real life (as reflected by the behavior ratings) may be why the endorsements in the present study show higher validities than the judgments for the Defensiveness subgroups. The fact that the behavior ratings of the Low Defensive group covaried with the judgments to a greater degree than did the behavior ratings of the High Defensive group would also provide an explanation for the judgment validities being higher for the Low than for the High Defensive *Ss*. This follows from the fact that there were few significant differences between the judgment scores of the two groups but there were relatively more significant differences between their behavior rating scores.

The finding that High Defensive individuals present favorable impressions in real life poses an interesting theoretical question. Do these individuals create favorable impressions intentionally or are they unaware of creating them? If they are unaware of the way in which they present themselves, that is, if favorable responding is part of their "life style" (cf. Crowne & Marlowe, 1964), then it is clear that personality scales are reflecting, in part, this "life style" of positive presentation. On the other hand, if they are aware of playing this favorable role, should one accept their scale scores or should one be skeptical and begin a search for their "real" or "true" personalities? If one commits himself to the definition of behavior of primary research interest as that which is endorsed on structured tests, as the present researcher has done, then one must accept by this definition all reliable test responses as being indicators of "true" personality in the sense that each response is determined by certain personality characteristics. In this sense one would accept the endorsement validity correlations of the High Defensive *Ss* as they stand. One would also be obliged to account for all reliable variance obtained under the judgment sets, including the relatively modest portion which proved to be related to criterion trait ratings. In this latter case, it is believed that the correlations between the judgments and the trait ratings were small in relation to the endorsement validities, because the judgment tasks differ in nature from the endorsements. The judgment tasks, being indirect, did not require *Ss* to reveal their personality traits in the same sense as did the endorsements. The judgments asked for an opinion of how desirable an item is in oneself or in others, or how frequently others would endorse the item.

One implication of the present study, therefore, is that endorsement rather than judgment instructions be used if one desires a valid measure of personality by means of structured scales and trait-rating criteria. The relatively high criterion validities, especially in the case of High Defensive and High Social Desirability respondents, justify the use of endorsements. The empirical support for the uniqueness of judgments suggests that more work needs to be done in uncovering the properties of judgments before applying them to the assessment of personality traits. The unique properties of the individual judgment sets were revealed in the present study by the interactive effects obtained between judgments and traits. For example, under one judgment set the validity coefficient is high while under another judgment set the validity coefficient for the same trait is low. Similar interactive effects between judgments and scale content are reported by Jackson and Singer (1967). In their study *Ss* were found to attribute the more pathological traits to others to a greater degree than to themselves. Further support for the complexity of judgments comes from a study by Jackson (1964b). Jackson's results suggested that in the case of certain traits a complementary-needs hypothesis may be warranted. For example, *Ss* scoring high on the trait Dominance under the endorsement condition were found to judge Submissiveness as being desirable in

others. The same held true for the complementary traits Succorance and Nurturance. Findings such as these reveal some of the problems involved in using judgments for assessing personality traits. It appears that a fruitful approach would be to examine each type of judgment set separately in relation to specific item content. In this way the complementarity-of-needs hypothesis as well as a similarity-of-needs hypothesis could be more precisely studied and the effects of different judgment sets on different traits more clearly delineated.

REFERENCES

CAMPBELL, D. T. The indirect assessment of social attitudes. *Psychological Bulletin*, 1950, **47**, 15–38.

CROWNE, D. P., & MARLOWE, D. *The approval motive*. New York: Wiley, 1964.

EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule*. New York: Psychological Corporation, 1954.

GOLDBERG, L. R., & RORER, L. G. Use of two different response modes and repeated testings to predict social conformity. *Journal of Personality and Social Psychology*, 1966, **3**, 28–37.

JACKSON, D. N. Assessing conformity with desirability judgments. *American Psychologist*, 1961, **16**, 446. (Abstract)

JACKSON, D. N. Desirability judgments as a method of personality assessment. *Educational and Psychological Measurement*, 1964, **24**, 223–238. (a)

JACKSON, D. N. *Personality Research Form*. Form A. Author, 1964. (b)

JACKSON, D. N., & MESSICK, S. *Differential Personality Inventory*. Pennsylvania State University: Authors, 1964.

JACKSON, D. N., & SINGER, J. E. Judgments, items, and personality. *Journal of Experimental Research in Personality*, 1967, **2**, 70–79.

LOOMIS, R., & SPILKA, B. Social desirability and conformity in a test situation. *American Psychologist*, 1963, **18**, 358. (Abstract)

MESSICK, S. Dimensions of social desirability. *Journal of Consulting Psychology*, 1960, **24**, 279–287.

ROSEN, E. Self-appraisal, personal desirability, and perceived social desirability of personality traits. *Journal of Abnormal and Social Psychology*, 1956, **52**, 151–158.

STRICKER, L. J., MESSICK, S., & JACKSON, D. N. Judgmental consistencies as predictors of social behavior. Paper presented at the meeting of the American Psychological Association, New York. September 1966.

(Received July 5, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 3, 234–239

EFFECTS OF THE TYPE OF STIMULUS EMPLOYED AND THE LEVEL OF SUBJECT AWARENESS ON THE DETECTION OF DECEPTION [1]

RICHARD I. THACKRAY [2] AND MARTIN T. ORNE [3]

Institute of the Pennsylvania Hospital and University of Pennsylvania

An experiment was designed to compare detection of personal material with that of neutral material made relevant only in the experimental context. Each S acted the role of an espionage agent and attempted to conceal both his personal identity and certain code words he had learned. Personally relevant material was found to be significantly more detectable than the neutral material, although consequences of detection were the same for both types. In addition, the study attempted to provide exploratory data concerning physiological responsivity to lie-detection stimuli when Ss were unaware that responses were being monitored. While there was no evidence that detection was inferior under the nonaware condition, difficulties in achieving a completely convincing nonaware situation suggest caution in generalizing from these findings.

Previous studies in detection of deception have reported varying rates of detection which appear to be related to the type or nature of the stimulus situation employed. Lykken (1960), for example, obtained 100% accuracy in matching idiosyncratic material (mother's name, father's name, etc.) to a particular *S*. Somewhat lower detection rates, ranging from 94% (Lykken, 1959; Marston, 1921) to 50% (Landis & Wiley, 1926), are reported when mock crimes are employed. Still lower rates, ranging from 83% (Ruckmick, 1938) to 40% (Landis & Wiley, 1926), are found with such "neutral" stimuli as numbers, letters, etc. While these studies would suggest that the greater the degree of affect associated with the type of stimuli employed the greater the obtained detection rate, differences between studies in methods of stimulus presentation, analysis of data, particular physiological measures used, etc., make valid comparisons virtually impossible.

Since systematic comparison of different types of stimulus material within a single study has not previously been reported, a primary aspect of the present study was to explore the importance of the stimulus dimension as a determinant of detection rate. Specifically, detection rates derived from a mock crime type of paradigm similar to one previously used (Thackray & Orne, 1967) were compared with rates obtained from personal or idiosyncratic material.

The study was also designed to provide some exploratory data about *S*'s physiological responsivity to lie-detection stimuli when unaware that his responses are being monitored. A number of studies have suggested that detection of deception is difficult or impossible if *S* is not motivated to deceive or if there are no consequences to being detected (Chappell, 1929; Gustafson & Orne, 1963). Thus, by eliminating *S*'s awareness of being monitored, he might not at that time be attempting to deceive, would therefore have nothing to hide, and would be difficult to detect. Under nonaware conditions, a difference between personally relevant words and neutral words which achieve relevance only in the context of the experiment might be particularly striking. For example, hearing one's own name is likely to result in a greater

[1] This work was supported in part by the United States Army Research and Development Command, Department of the Army, under Research Contract No. DA-49-193-MD-2647.

[2] Now at the Civil Aeromedical Research Institute, Federal Aviation Administration, Oklahoma City, Oklahoma.

[3] The authors wish to express their appreciation to Santina Clauser and Luise Miller for their assistance in this experiment and in particular to William Orchard, who was E_1, and to Mary-Jo Bryan and George Smiltens, who served as technicians. They also wish to thank Frederick J. Evans, Ulric Neisser, Donald N. O'Connell, Emily C. Orne, and Peter W. Sheehan for their helpful comments during the preparation of the manuscript.

physiological response than hearing other names independent of the detection-of-deception situation. On the other hand, when deception is not being attempted, such differentially greater physiological responsivity might not occur with material that is relevant only in the context of the experiment.

The physiological response measure employed in the present study was the galvanic skin response (GSR), since previous laboratory studies have shown it to be the single best index of deception (Ellson, Davis, Saltzman, & Burke, 1952; Kubis, 1962; Kugelmas & Lieblich, 1966; Violante & Ross, 1964). To use this measure under a nonaware condition, however, presents certain difficulties. There appear to be no completely adequate techniques for insuring uniform and continuous electrode contact without having electrodes physically attached to *S*. It was felt, though, that a situation might be devised in which electrodes were attached, but in which *S* would believe that no recordings were actually being taken. The situation chosen approximated closely the procedure followed in actual field polygraph examinations. In the field procedure, each question is carefully reviewed with the suspect prior to the actual examination. In the present experiment, GSRs were surreptitiously recorded during this review phase in the hopes of approximating a nonaware condition.

METHOD

Subjects

Twenty-eight male undergraduate college students were recruited from local universities. All were volunteers and none had previously participated in detection-of-deception experiments.

Procedure

The critical stimulus material consisted of three common words and three personally relevant items of information: *S*'s first name, last name, and date of birth. Each critical stimulus was embedded among five similar stimuli. Prior to *S*'s arrival, an assistant randomly selected one card from each of three sets of six cards containing names of common birds, trees, and colors, respectively, which would subsequently serve as "code words" to be learned by a given *S*.[4] He also selected five last names from a large list of names of similar ethnic origin, five first names from a list of common male names, and five random dates (month and day). The *S*'s own last name, first name, and date of birth were then randomly positioned in their respective sets. The only restriction placed on the sets was that the first item served as a dummy in the interrogation and consequently could never be a significant stimulus.

Each *S* was seen by E_1 who explained that the effectiveness of the polygraph as a lie detector was being studied. He was told that he was to play the role of an espionage agent who knows certain critical code words and that he would be given a polygraph examination subsequently by the interrogator (E_2). It was further explained that, due to a security leak, the interrogator was in possession of a pool of code words as well as the names and birthdates of six agents. His own name and the code words to be learned would be contained in this list. To maintain his anonymity, however, he would be introduced as "*S* 27" and the interrogator would not know whether he (the *S*) was one of the individuals whose names he had or whether he was in fact innocent (belonged to a control group who did not receive such information). The *S*'s task was to convince the interrogator that he was innocent and that none of the code words or personal information had any meaning to him. It was explained to him that while it was difficult to deceive the polygraph, some highly intelligent and mature individuals had been able to do so. The *S* was then given a booklet which was constructed to maximize overlearning of the three code words.

After *S* had completed the initial phase of the experiment, he was introduced to E_2 as *S* 27. The E_1 simultaneously gave E_2 the set of stimulus cards specific for the *S*. While no innocent *S*s were used in this study, E_2 was actually blind as to the relevant code words and personal information concerning the *S*. As *S* entered the room, the polygraph was in the process of being "calibrated." The technician completed her task and turned off the instruments as *S* was being seated in a stratolounger chair facing away from the equipment.[5] The E_2 treated *S* analogously to the way an interrogator might. He explained the basic operation of the polygraph and stressed the virtual infallibility of the instrument. He explained to *S* that if he were guilty the equipment would certainly register this fact, and he could save considerable time and trouble by confessing (no one did). The interrogator explained that he would go over each of the questions with *S* to make certain there was no ambiguity, surprise, or confusion regarding any of the questions or the procedure itself. The *S* was told that after the familiarization phase, the polygraph equipment

[4] The actual stimuli employed were as follows: Colors—PURPLE, ORANGE, BROWN, GRAY, GREEN, BLUE; Birds—CROW, PIGEON, OWL, WREN, ROBIN, SPARROW; Trees—OAK, HEMLOCK, ELM, SPRUCE, MAPLE, PINE.

[5] The "calibration" was to acquaint *S* with the sound of the polygraph and to reinforce his belief that no recording was being attempted prior to the actual examination.

would be turned on, adjusted, and the examination itself carried out.

Electrodes and transducers for recording GSR and six other variables were then attached. (The remaining variables will be discussed in a separate paper.) The leads from the various electrodes, transducers, and associated equipment were not connected to the polygraph and were left loosely arranged on *S*'s lap. One of these polygraph leads was actually a dummy, however, and led from one of several small transducer boxes located on *S*'s chair. The GSR electrodes led to this box which contained the transmitter and bridge circuit of an FM telemetry system.[6] The GSR was telemetered to a separate room and recorded on a separate polygraph elsewhere. It was explained to *S* that the electrodes had to be attached at this time in order to allow sufficient time for them to "stabilize" prior to the later polygraph examination.

After all electrodes were attached to the *S*, E_2 informed him that the questions would now be reviewed with him. Since the purpose of the preinterrogation was to familiarize him completely with all aspects of the polygraph examination which would subsequently follow, *S* was told not only to listen to the questions, but to answer each with a "yes" or "no" just as he would later during the examination. The *S* was then questioned on each word in the six stimulus sets. The items were presented by E_2 as informally as possible, allowing approximately 6 sec. between each item to provide sufficient time for the telemetered GSRs to be recorded. Half the *S*s received the code-word material first followed by the personal material, while the order was reversed for the remaining half.

Upon completion of the preinterrogation phase, the various leads were connected to the polygraph and the equipment was adjusted by an assistant acting as the polygraph operator. The *S* was then informed that the examination would begin. Questions were asked in the same order employed in the preinterrogation, and *S* was again asked to answer each with a "yes" or "no." A 16-sec. interstimulus interval was used to allow for differing latencies in the physiological response measures employed. Following completion of the questions, *S* was disconnected from the polygraph and taken to E_1 for a postexperimental inquiry.

The data consisted of the telemetered GSRs obtained during the preinterrogation as well as during the "actual" polygraph examination. Since the telemetered data did not yield absolute values of resistance, the magnitude of response was obtained by measuring in millimeters the maximum pen deflection to occur within 5 sec. following the onset of each verbal stimulus. The response to the critical stimulus was ranked against the four control words in its respective set, and a rank of 1 was assigned to the largest response. Mean ranks were then determined for each *S* on both the code words and the personal items.

[6] Onyx Designs, Inc.

RESULTS

Table 1 displays the mean ranks obtained for both types of stimulus material under the nonaware and aware conditions. A mean rank of 1 would indicate perfect detection; it is noteworthy that in every comparison personal material yields a smaller mean rank than code-word material. To test the possibility of combining orders, comparisons were made between ranks of personal words under both orders of presentation, and likewise between code words under both orders. Separate comparisons were made for the nonaware and aware series. None of the differences was significant at the .05 level (Mann-Whitney U; two-tailed). The orders were then combined and the results shown in Table 2. As indicated in this table, Wilcoxon matched-pairs signed-rank tests revealed a significant superiority of personal words over code words.

In order to determine whether detection rate differed as a function of awareness, chi-square tests were separately performed on the personal- and code-word distributions. Since the expected chance value of the mean ranks is 3, the number of mean ranks above and

TABLE 1

MEAN GSR RANKS ON CODE WORDS (CW) AND PERSONAL WORDS (PW) FOR BOTH ORDERS OF PRESENTATION UNDER BOTH NONAWARE AND AWARE CONDITIONS

Condition	Order	Stimulus material	$\bar{X}$ rank
Nonaware	PW–CW[a]	PW	1.38
		CW	1.67
	CW–PW[b]	CW	1.63
		PW	1.23
Aware	PW–CW	PW	1.46
		CW	1.70
	CW–PW	CW	1.55
		PW	1.42

[a] $N = 13$.
[b] $N = 15$.

TABLE 2

MEAN GSR RANKS ON COMBINED CODE WORDS (CW) AND PERSONAL WORDS (PW) UNDER NONAWARE AND AWARE CONDITIONS

Condition	Stimulus material	$\bar{X}$ rank	x^2	T
Nonaware	PW	1.30	26.03***	26.5**
	CW	1.65	22.32***	
Aware	PW	1.44	21.33***	38.5*
	CW	1.62	26.03***	

Note.—$N = 28$.
* $p < .025$, one-tailed test.
** $p < .01$, one-tailed test.
*** $p < .001$.

below this value was determined for each distribution and evaluated against the expected frequencies. As Table 2 indicates, the four chi-square values were significant, indicating better than chance detection under both awareness conditions. To test for any relative superiority of the two conditions, a comparison was made between ranks of personal words in the nonaware and aware series. Likewise, a similar comparison was made between ranks of code words in both series. Wilcoxon tests yielded nonsignificant z values of 1.09 and .09, respectively.

During the postexperimental inquiry each *S* rated how detectable he thought he was on the three personal-word and three code-word items. Mean ranks for these two types of stimulus material were determined for each *S*, and the difference between the distributions tested with a sign test. Results indicated *S*s felt themselves to be more detectable on the personal-word material ($p < .001$).

DISCUSSION

Previous research has focused on the consequences of detection and *S*'s motivation to deceive as significant variables affecting the likelihood of differentially increased physiological responses (Gustafson & Orne, 1963, 1965). In the present study, responding to either personal items or code words would lead to detection. The *S* was aware that a response to either type of stimulus would have the same consequences; that is, he would reveal himself to be a guilty person. Yet, the results indicate that detection of personally relevant material is superior to relatively neutral stimuli which are made relevant only in the context of the experiment. If *S*'s motivation to deceive and the consequences of deception are held constant, however, it seems reasonable that personal material would be more likely than code material to evoke some form of emotional response. These findings are not inconsistent with previously reported results that situational factors have marked effect on detection rates, since significant detection was achieved with neutral stimuli made relevant in the experimental context. Thus, one would expect a decrease in detection rates with both kinds of stimuli in a situation where motivation is markedly reduced. The findings do suggest, though, that situational determinants alone are not sufficient to account for the obtained detection rates. If detection were only a function of its consequences, personal material would not have been detected better than code-word material.

The differential response between personal-word and code-word material observed in the present study is consistent with the findings of Oswald (1962, pp. 47–50) that *S*s during sleep are more likely to rouse to their own name than to other stimuli presented at equal volume even if motivating contingencies are associated with other stimuli, though motivating contingencies do affect the probability of arousal to previously neutral stimuli during sleep. Whether one wishes to conceptualize the observable difference of response to highly relevant personal material (such as one's name) as due to past differential affective conditioning or to differences in meaning is largely a matter of semantic preference. That personally relevant material tends to evoke a differential affective response has also been shown in quite a different context in the work of Wolff (1943). He found that individuals gave a greater emotional reaction to their own forms of self-expression even when these were not recognized as such. The detection-of-deception paradigm used here is one where *S*'s physiological response to a particular stimulus is compared with that to other similar stimuli. In past research (Gustafson & Orne, 1963),

it has been shown that previously neutral stimuli can be made relevant to the experimental situation and yield differentially greater physiological responses. The previous work by both Oswald and Wolff has shown that personally relevant stimuli yield greater affective responses than neutral stimuli in a situation other than a detection-of-deception experiment. The authors of this study would interpret their findings to suggest that the two sources of increased physiological responsivity, situational and preexisting personal relevance, may each independently contribute to detection.

The *S*s rated themselves as more detectable on personal than on code material in the postexperimental inquiry. A clear interpretation of this finding is difficult. They may have perceived themselves as responding to personal items during the polygraph examination or they may merely have expected that personal material should be more detectable than code words. Further research is needed to clarify the results.

In designing the present study, it was hoped that the nonaware condition would allow a comparison between personal- and code-word stimuli in a situation where, as far as the *S* knew, a physiological response would not have the consequence of detection. Under these circumstances one might expect that the code words would fail to yield a greater physiological response than their appropriate control stimuli, whereas this might not be the case with idiosyncratic or personal material. This difference in relative response to the two classes of stimuli might become smaller or disappear entirely under the aware condition of interrogation, since at that time the *S* would be motivated to attempt to conceal his physiological responses to both code words and personally relevant words.

The actual findings indicated no significant differences between these two conditions. Personally relevant words yielded differentially greater physiological response than code words in both conditions; however, code words themselves were identified as readily in their nonaware condition as in their aware condition. It is of course possible that *S*s did not believe appearances and thought that somehow they might still be monitored (as indeed they were) during the nonaware condition. Not only was every possible precaution taken to minimize suspicion, but very careful postexperimental inquiry was carried out on this issue. Only one *S* appeared to have any suspicion in this regard; most *S*s, however, verbalized feelings of discomfort during the nonaware condition. They indicated they felt strange having all the electrodes attached to them and anticipating the forthcoming polygraph examination. Some *S*s even indicated that if they had been recorded during this initial period detection probably would have been easier than during the actual examination. It would seem that having a great many electrodes attached and expecting the forthcoming polygraph examination created a condition for most of the *S*s that was similar to that of the actual interrogation. It should be emphasized that *S*s did attempt to deceive *E* during the preliminary part of the examination. Further work will be needed to clarify whether the augmented GSR associated with the code words under the nonaware condition was a function of the *S*'s motivation to deceive *E* independent of his belief that records were being obtained or whether it was related to the number of electrodes that were attached to *S* and the special situation in which he found himself. A definitive answer about the effect of the awareness of being monitored on physiological responsivity will require considerably more investigation, and great care should be taken in generalizing from the findings reported here.

REFERENCES

CHAPPELL, M. N. Blood pressure changes in deception. *Archives of Psychology, New York,* 1929, **17**, 1–39.

ELLSON, D. G., DAVIS, R. C., SALTZMAN, I. J., & BURKE, C. J. A report of research on detection of deception. (Contract N6onr-18011 with Office of Naval Research.) Distributed by Department of Psychology, Indiana University, 1952.

GUSTAFSON, L. A., & ORNE, M. T. Effects of heightened motivation on the detection of deception. *Journal of Applied Psychology,* 1963, **47**, 408–411.

GUSTAFSON, L. A., & ORNE, M. T. Effects of perceived role and role success on the detection of deception. *Journal of Applied Psychology,* 1965, **49**, 412–417.

KUBIS, J. F. Studies in lie detection: Computer feasibility considerations. Contract AF30(602)-2270, Project 5534. Prepared for Rome Air Development Center, TR62-205. Distributed by Arlington, Virginia: Armed Services Technical Information Agency, AD-284902, June 1962.

KUGELMASS, S., & LIEBLICH, I. Effects of realistic stress and procedural interference in experimental lie detection. *Journal of Applied Psychology*, 1966, **50**, 211–216.

LANDIS, C., & WILEY, L. E. Changes of blood pressure and respiration during deception. *Journal of Comparative Psychology*, 1926, **6**, 1–19.

LYKKEN, D. T. The GSR in the detection of guilt. *Journal of Applied Psychology*, 1959, **43**, 385–388.

LYKKEN, D. T. The validity of the guilty knowledge technique: The effects of faking. *Journal of Applied Psychology*, 1960, **44**, 258–262.

MARSTON, W. M. Psychological possibilities in the deception test. *Journal of Criminal Law and Criminology*, 1921, **11**, 551–570.

OSWALD, I. *Sleeping and waking*. Amsterdam: Elsevier, 1962.

RUCKMICK, C. A. The truth about the lie detector. *Journal of Applied Psychology*, 1938, **22**, 50–58.

THACKRAY, R. I., & ORNE, M. T. Methodological studies in detection of deception. United States Army Medical Research and Development Command Research Report filed with Defense Documentation Center No. AD-645102, January 1967.

VIOLANTE, R., & ROSS, S. A. Research on interrogation procedures. Contract Nonr-4129(00), Stanford Research Institute, Menlo Park, California, 15 May—31 October, 1964.

WOLFF, W. *The expression of personality*. New York: Harper, 1943.

(Received July 12, 1967)

List of Manuscripts Accepted for the *Journal of Applied Psychology*

Brand Awareness: Effects of Fittingness, Meaningfulness, and Product Utility: Rabindra N. Kanungo*: Department of Psychology, Dalhousie University, Halifax, Nova Scotia.

Age, Cigarette Smoking, and Tests of Physical Fitness: Kenneth E. David*: University of Maryland, Far East Division, APO San Francisco 96525.

Relationship between Perceptual Style and Simulator Sickness: Gerald V. Barrett* and Carl L. Thornton: Director, Human Factors Laboratory, Life Sciences Research Department, Goodyear Aerospace Corporation, 1210 Massillon Road, Akron, Ohio 44315.

Job Satisfaction among Graduate Students' Intrinsic versus Extrinsic Variables: Edward L. Levine* and Joseph Weitz: 720 East 31st Street, Apartment 6H, Brooklyn, New York 11210.

Heuristic Simulation of Psychological Decision Processes: Robert D. Smith*: Department of Management, 120 Boucke Building, The Pennsylvania State University, University Park, Pennsylvania 16802.

Comparison of Questionnaire Responses with and without Preceding Interviews: Clayton P. Alderfer*: Graduate School of Business and Public Administration, Cornell University, Ithaca, New York 14850.

Effect of Control Placement on Information Transfer Rate Using Bisensory Signals: A. D. Perriment*: Department of Psychology, Monash University, Clayton, Victoria, Australia.

A Psychophysical Determination of Equitable Payment: A Methodological Study: Sheldon Zedeck* and Patricia Cain Smith: Department of Psychology, Bowling Green University, Bowling Green, Ohio 43402.

Task Coherence, Training Time, and Retention Interval Effects on Skill Retention: James C. Naylor, George E. Briggs,* and Walter G. Reed: The Human Performance Center, Ohio State University, 404-B West 17th Avenue, Columbus, Ohio 43210.

Use of Importance as a Weighting Component of Job Satisfaction: Patricia Smith Mikes and Charles L. Hulin*: Department of Psychology, University of Illinois, Urbana, Illinois 61801.

Word Legibility as a Function of Letter Legibility, with Word Size, Word Familiarity, and Resolution as Parameters: R. L. Erdmann* and A. S. Neal: Department 924, IBM Corporation, P.O. Box 66, Los Gatos, California 95030.

Evaluating Mortgage Loan Risk: William D. Buel*: The Vernon Psychological Laboratory, 221 North LaSalle Street, Chicago, Illinois 60601.

---

* Asterisk indicates author for whom the address is snpplied.

(Continued on page 262)

Journal of Applied Psychology
1968, Vol. 52, No. 3, 240–244

OCCUPATIONAL VALIDITY OF THE GENERAL APTITUDE TEST BATTERY

STEPHEN E. BEMIS

United States Employment Service, Washington, D. C.

20 yr. of occupational validation on the General Aptitude Test Battery (GATB) are summarized in terms of average validity with (a) job proficiency and training criteria and (b) concurrent and longitudinal studies. These 4 categories of studies are investigated from the standpoint of both single and multiaptitude prediction. The relative efficiency of many of the 9 GATB aptitudes in predicting criteria is dependent on whether job proficiency or training criteria are used. Longitudinal studies tend to have higher-aptitude criterion correlations than concurrent studies. The median validity of batteries of GATB aptitudes tends to be higher with studies using training criteria or the longitudinal design than the validity of batteries using job-proficiency criteria or the concurrent design. The summary is based upon 424 studies involving over 25,000 employees, applicants, trainees, and students.

In his recent book, Ghiselli (1966) summarizes the success of researchers in predicting occupational performance with aptitude tests. Ghiselli's book provided the impetus for this article which summarizes the United States Employment Service (USES) research results in terms of the average validity of each General Aptitude Test Battery (GATB) aptitude against job proficiency and training criteria and in longitudinal and concurrent validation studies. The results obtained when batteries of GATB aptitudes are developed with each of these two types of criteria and two types of designs are also compared. The summary is based on 424 studies involving over 25,000 employees, applicants, trainees, and students.

Background

The USES test research program, initiated in 1935, led to the development of the USES General Aptitude Test Battery (GATB) published in 1947. Occupational test development analysts are studying aptitudinal requirements, as shown by the GATB, for successful on-the-job performance in 36 state employment services affiliated with the USES. To date, GATB validity studies have been conducted on over 500 occupations. Although USES aptitude test research has taken several new directions in recent years (Droege, 1966; Dvorak, 1965; Dvorak, Droege, & Seiler, 1965), such as the development of literacy skills and nonreading aptitude tests, and achievement and proficiency tests are being developed, the validation of the GATB for specific occupations is still the major component of the USES test research program.

The nine aptitudes measured by the GATB are General Learning Ability (G), Verbal Aptitude (V), Numerical Aptitude (N), Spatial Aptitude (S), Form Perception (P), Clerical Perception (Q), Motor Coordination (K), Finger Dexterity (F), and Manual Dexterity (M). Each aptitude is scaled to a standardized mean of 100 and standard deviation of 20 based on a general working population sample (United States Department of Labor, 1967).

Each of the research studies summarized below resulted in a specific occupational test battery composed of minimum cutting scores on two–four GATB aptitudes. These test batteries are currently in use by approximately 1,600 local offices of the state employment services. (Information on each of these studies as well as a historical and technical description of the GATB are available elsewhere [United States Department of Labor, 1967].) Data on: (*a*) older studies which have been replaced by new validations, (*b*) studies which did not yield norms, and (*c*) studies which yielded norms considered suitable for single plant use only (because of small sample size) are not included in this summary. (It is estimated that 10% of USES studies fail to yield norms.) While USES studies have employed numerous types of criteria, ratings

by supervisors or instructors have been used most frequently.

PROCEDURE AND RESULTS

Job Proficiency versus Training Criteria

The median correlation was computed for each aptitude for the 74 USES studies in which a training criterion had been used. Similarly, the median correlation was computed for each aptitude in a randomly selected 20% sample of the 345 USES studies in which a job-proficiency criterion was used. Grand medians were computed for each type of criterion. These results are shown in Table 1.

Table 1 shows that the typical aptitude-criterion correlation obtained in USES studies is approximately .22 regardless of the type of criterion used. However, Table 1 also shows that the cognitive aptitudes (G, V, N, S) typically correlate to a higher degree with training criteria than they do with job-proficiency criteria. The additional amount of training variance accounted for is greatest for Aptitude G (approximately 9%) and least for Aptitude S (approximately 3%), with Aptitudes V and N showing about the same additional amount (approximately 5%). On the other hand, the manipulative aptitudes (K, F, M) tend to correlate lower with training criteria than with job-proficiency criteria. These correlations reflect differences in the amount of criterion variance accounted for of $2\frac{1}{4}$% (Aptitude F) and $3\frac{1}{3}$% (Aptitudes K and M) in favor of the job proficiency. The perceptual aptitudes (P and Q) tend to have above the same level of validity with either job-proficiency or training criteria. (It should be noted that all GATB aptitudes are typically included in USES research regardless of the type of criterion.)

Concurrent versus Longitudinal Validity

Ghiselli's calculation of the average validities obtained with job-proficiency and training criteria was only an aside in his attempt to summarize the predictive power of different categories of aptitude tests and the predictability of different categories of jobs. However, it would have been interesting if he had calculated the average validity of another classification of studies—concurrent versus longitudinal. Longitudinal validation is usually held to be preferable from the standpoint of experimental design because it yields "predictive validity," but because of practical considerations much of the USES research and much of the occupational validation done elsewhere is on a concurrent basis.

The concurrent–longitudinal validity comparison is related to the job-proficiency–training criteria comparison because in USES research many studies utilizing training criteria are longitudinal. However, studies utilizing job-proficiency criteria are both concurrent and longitudinal. The median correlation was computed for each aptitude for the 72 USES studies conducted according to a longitudinal research design. Similarly, the median correlation was computed for each aptitude for a randomly selected 20% sample of the 353 USES studies conducted according to the concurrent research design. These results and the associated grand medians are reported in Table 2.

Table 2 shows that the typical aptitude-criterion correlation obtained in USES longitudinal studies is approximately .22 while the

TABLE 1

MEDIAN VALIDITIES OF GATB APTITUDES

Type of criterion	No. studies	Aptitude									Grand *Mdn*
		G	V	N	S	P	Q	K	F	M	
Job proficiency	64–70[a,b]	25	19	21	17	23	20	23	22	23	22
Training	67–71[a]	39	30	31	25	22	22	14	16	14	22

[a] *N*s differ slightly by aptitude.
[b] 20% sample of all studies using a job-proficiency criterion.

TABLE 2

MEDIAN VALIDITIES OF GATB APTITUDES

Design of study	No. studies	Aptitude									Grand *Mdn*
		G	V	N	S	P	Q	K	F	M	
Concurrent	66–69[a,b]	27	22	24	18	15	21	16	19	20	20
Longitudinal	71–72[a]	34	22	28	26	27	22	19	21	22	22

[a] *N*s differ slightly by aptitude.
[b] 20% sample of all concurrent studies.

typical correlation is .20 in concurrent studies. Eight of the nine aptitudes also show slightly higher validities with the longitudinal design, while there is no difference on Aptitude V. The largest differences in Table 2 indicate that an additional 5% (Aptitude P), 3½% (Aptitude S), and 4¼% (Aptitude G) of the criterion variance is predicted in longitudinal studies than is predicted in concurrent studies. The slightly higher validities obtained in longitudinal studies are not due to a greater range of talent in this group as the medians of the aptitude standard deviations in the studies shown in Table 2 are either the same (S, K, F) or 1 point higher for the concurrent studies. The grand median standard deviations are .17 for concurrent studies and .16 for longitudinal studies.

Predictor Combinations

One consideration which Ghiselli omitted was the average validity obtained when several predictors are used. In the typical study, the multiple correlation technique is used to determine the best possible weighting of predictors. As pointed out by Dvorak (1956), the USES considers the multiple cutoff method to be superior to the multiple correlation method for USES occupational validation. Since that time USES has begun using the phi coefficient rather than the tetrachoric correlation to report the relationship between the variable dichotomies ("high" or "low" on the criterion versus "pass" or "fail" on a battery consisting of minimum qualifying scores on two–four aptitudes).

Table 3 shows the median phi coefficients obtained in USES studies when divided into the following categories: job proficiency and training, and concurrent and longitudinal.

The difference between the median phi coefficient for the concurrent type of study (.40) and the longitudinal type of study (.45) is greater than the difference between job-proficiency criterion studies (.40) and training criterion studies (.42). These results are generally consistent with those shown in Tables 1 and 2.

Additional Analyses

Two preliminary analyses were performed in an attempt to relate USES results to differences between studies with job proficiency and training criteria reported by Ghiselli.

Ghiselli reports grand average validities of .19 with job-proficiency criteria and .30 with training criteria. Ghiselli's figures appear to be based upon average correlations obtained in 107 matched sets of studies in which both job-proficiency and training criteria had been collected. The few USES batteries in which both job-proficiency and training criteria had been collected on the same occupation (11 matched sets) were similarly summarized. The results showed a median correlation of .24 for all aptitudes with job-proficiency criteria (97 correlations) and .26 with training cri-

TABLE 3

MEDIAN VALIDITY OF BATTERIES OF GATB APTITUDES

Type of study	*Mdn* phi coefficient
Job proficiency (N = 345)	.40
Training (N = 74)	.42
Concurrent (N = 353)	.40
Longitudinal (N = 71)	.45

Note.—Five studies in which a combined job-proficiency–training criterion was used were eliminated from the job-proficiency–training comparison but included in the concurrent–longitudinal comparison.

teria (93 correlations). Although the difference between the two median validities is slight, it is in the same direction as that reported by Ghiselli.

Before all of the data upon which the medians reported in Table 1 and 2 were readily available, an exploratory study was conducted. In this study, the highest of the nine aptitude-criterion correlations obtained in each of 282 USES studies with a job-proficiency criterion and 48 studies with a training criterion (a total of 330 studies) was singled out and the median of these distributions was obtained as a method for comparing training and a job-proficiency criteria. The results showed a higher median validity for studies with a training criterion (.44) than for studies with a job-proficiency criterion (.38). A parallel study was conducted with these data to compare median correlations from concurrent and longitudinal studies. Similar results were obtained. (Longitudinal studies yielded the higher validity.)

CONCLUSIONS

Cognitive aptitudes (G, V, N, S) are more useful for predicting training criteria than for predicting job-proficiency criteria. Manipulative aptitudes (K, F, M) are more useful for predicting job-proficiency criteria. However, considering the grand median validity for all aptitudes, the same level of prediction (.22) is obtained regardless of the type of criterion. The difference in the types of criteria which the cognitive and manipulative aptitudes predict best points to the need for a multiple-hurdle criterion for many validation studies. Different aptitudes (or other predictors) may be correlated with success in the book-learning and the work-performance aspects of learning the same job. A multiple-hurdle criterion has been used in some USES studies where it was recognized that the job-performance criterion could not be considered the sole measure of occupational success.

Most aptitudes (especially G, S, and P) tend to yield better validity in longitudinal studies than in concurrent studies. This may mean that validity coefficients based upon concurrent studies tend to give underestimates of the predictive validity of a test used in selection, ignoring of course the possibility of validity shrinkage in cross-validation. This fact and the median aptitude standard deviations obtained in longitudinal and concurrent studies seem to cast some doubt on the inate superiority frequently assumed for longitudinal studies. If the validity reported for a longitudinal study is based upon the aptitude test battery–criterion relationship for only those people who complete the minimum training time and are on the job at criterion collection time, the range of talent in the research sample may be no different than that which would be found in a comparable, but completed sooner, concurrent study. Procedures are being instituted in USES research to try to identify individuals who do not remain in the research sample until criterion collection time because of inability to learn or perform the job. Such individuals will be included in the low criterion group or the final four-fold table used to reflect the validity of test batteries developed by USES.

The above criticism of the longitudinal design is not meant to belittle other advantages of this design such as: (*a*) elimination of possible length of service contamination in the criteria, (*b*) elimination of possibility of higher scores on a predictor due to information or tasks learned in conjunction with the job, and (*c*) similarity of research sample to applicant sample.

Plans are now being made for additional analyses of USES data which will be similar to Ghiselli's summary of aptitude-criterion correlations obtained with various categories of jobs. For these analyses jobs will be grouped according to the new USES occupational structure (United States Department of Labor, 1965) rather than the earlier USES structure or the special structure which were both used by Ghiselli.

Research along several lines must be intensified if the modest but useful validities reported above are to be raised. These lines are:

1. Methods of criterion improvement (about which much is said but little is done).
2. Use of moderator variables (recent work by Hobert & Dunnette, 1967, is encouraging).
3. Use of variables such as interest tests and biographical inventories in addition to

aptitude tests in the prediction equation (interest tests offer a possibility for a significant and early increase in validity coefficients only if a systematic and acceptable way of using interest tests in routine selection situations can be derived).

None of the three areas of research mentioned above is new or unexplored, but they have not received anywhere near the same attention as aptitude test research either by USES or other researchers. USES is now moving into a research program which will attempt to improve criteria and explore the use of moderator variables and nonaptitude predictors.

TECHNICAL NOTES

The median correlation is used to summarize the USES data rather than the weighted numerical average used by Ghiselli because of the computational problem involved in transforming phi coefficients for purposes of averaging. However, the weighted numerical average was computed for a number of subsamples of aptitude–criterion correlations and was found to be approximately the same as the median.

Although a few studies are based on samples of less than 50 individuals (older studies or cross-validation studies) and a few studies are based on samples of several hundred individuals, the majority of the studies (about 80%) are based on *N*s of 50–75. The boundaries of this range are dictated at the lower end by a long-standing USES requirement that norms must be based on samples of at least 50 in order to be used nationally and at the upper end by the practical problem of trying to find more than 75 people performing one specific job.

REFERENCES

DROEGE, R. C. GATB longitudinal maturation study. *Personnel and Guidance Journal*, 1966, **44**, 919–930.

DVORAK, B. J. Advantages of the multiple cut-off method. *Personnel Psychology*, 1956, **9**, 45–47.

DVORAK, B. J. Program research on effectiveness of USES tests. *Employment Service Review*, 1965, 2(10), 62–64.

DVORAK, B. J., DROEGE, R. C., & SEILER, J. New directions in U. S. Employment Service aptitude test research. *Personnel and Guidance Journal*, 1965, **44**, 136–141.

GHISELLI, E. E. *The validity of occupational aptitude tests*. New York: Wiley, 1966.

HOBERT, R., & DUNNETTE, M. D. Development of moderator variables to enhance the prediction of managerial effectiveness. *Journal of Applied Psychology*, 1967, **51**, 50–64.

UNITED STATES DEPARTMENT OF LABOR. *Dictionary of occupational titles*. Vol. 1 & 2. Washington, D. C.: Author, 1965.

UNITED STATES DEPARTMENT OF LABOR. *Manual for the General Aptitude Test Battery*. Section 3, Development. Washington, D. C.: Author, 1967.

(Received July 17, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 3, 245–253

SCALES FOR RATING THE TASTE OF WATER[1]

WILLIAM H. BRUVOLD

School of Public Health, University of California, Berkeley

Hedonic, quality, action tendency, and quality-action tendency rating scales, each designed to evaluate the taste of domestic water, were constructed. Near linear relations were found between equal appearing and successive interval scale values for the items comprising each scale. Average reliability coefficients for scales ranged from .62 to .77 in the laboratory work here reported. Mean scale ratings of sodium solutions were found to differ significantly by anion and by concentration level on each scale. In consumer survey research, the quality rating scale was found to correlate .83 with previously developed attitude scales, and to be significantly and linearly related to the total amount of common mineral present in a water supply. Use of the rating scales to develop limiting standards for minerals, and for other constituents commonly found in domestic water, was discussed.

A research project has been initiated (Bruvold, Ongerth, & Dillehay, 1967; Dillehay, Bruvold, & Siegel, 1967) which is intended to produce objective data that can be used to recommend limiting standards for common mineral content in domestic water. The desired standards are intended to insure the potability of water served to the public. Lack of research regarding the relationship between amount of common mineral content in water and potability has prevented the United States Public Health Service (1962), and responsible state agencies, from establishing definitive standards for these minerals in drinking water. The absence of definitive standards means that new public water supply permits may be issued for unpotable waters. This problem is particularly acute in rapidly growing population centers located in the southwestern United States. Also, the absence of the needed standards serves to impede the action necessary to provide all consumers with fully potable water for daily consumption.

The present report outlines the construction and evaluation of four rating scales that can be used to describe quantitatively the relation between common minerals, or other dissolved constituents, and the taste of water. Three of the scales were new, and one represented a refinement of a scale used in an earlier work (Bruvold & Pangborn, 1966). Specifically, the present paper has three major aims. The first is to describe procedures used to construct the four scales. The second is to report results from a laboratory study designed to evaluate the four scales. The third is to report results of consumer surveys designed to evaluate the quality rating scale.

Scale Construction

Combination Scale

A rating scale constructed by rank ordering statements used by *Ss* to describe the taste of water in a free response setting was employed in initial laboratory research (Bruvold & Pangborn, 1966). This rating scale contained 14 items, each item consisting of a quality statement in combination with an action tendency statement. These items were the first to be scaled.

Method. The 14 combination items were randomly arranged on a rating form with 81 additional items describing the taste of water in general terms. All 95 items were presented as the third part of a mimeographed booklet. The first part of the booklet contained 105 adjectives evaluating the taste of water, while the second part contained 73 statements generally evaluating water quality. The three sets of items, minus the combination statements dealt with here, formed the item pools from which three at-

[1] This research was supported in part by funds provided by the United States Department of the Interior, Office of Water Resources Research, as authorized under the Water Resources Research Act of 1964, and by the University of California Water Resources Center; and in part by United States Public Health Service, National Institutes of Health, General Research Support Awards 5-S01-FR-5441-05 to the School of Public Health, University of California, Berkeley and 1-S01-FR-5549-04 to the California State Department of Public Health.

titude scales have been constructed (Dillehay et al., 1967).

The items were judged by 44 *S*s (Dillehay et al., 1967). Judgments were made on the usual 11-category scale employed in the method of equal appearing intervals; standard instructions associated with this method were used (Edwards, 1957). All *S*s judged all items in the same order.

Results. The median category assignments and *Q* values were determined for each combination item using standard analytical procedures (Edwards, 1957). Also, means and standard deviations for each item were determined by computer using the iterative least-squares method of successive intervals developed by Diederich, Messick, and Tucker (1957). Weights of 2, 1, and 0 were employed in the latter analysis with cutoff points of ±0.7 and ±1.4. The relationship, over all items, for equal appearing and successive interval values, was very nearly linear.

Nine items were selected from the total group so that scale distances between successive items were approximately equal. Items forming the combination scale are shown in Table 1; scale and dispersion values are listed in the last two columns of Table 2.

New Scales

Three new rating scales were constructed after the work on the combination scale had been completed. One scale contained hedonic statements, one quality statements, and one action tendency statements. The three scales constructed were analogous to those described by Schutz (1965) for use with food products. The procedures employed to rate statements were those suggested by Torgerson (1958) in a discussion of subjective estimate methods.

Method. All items were phrased to refer directly to water. Nine hedonic items were constructed by rephrasing items from the scale used by Schutz (1965); 23 new hedonic items were also constructed. Thirty-four quality and 18 action tendency items were constructed using the same free responses to water samples employed to develop the original combination scale. An attempt was made to cover systematically the entire range of reaction to the taste of water in preparing each group of items.

All items from each group were typed, one to a card on small plain white cards and labeled, at random, with two digit numbers. The three groups of cards were presented, in a random order, to 53 *S*s, all of whom were employees of the California State Department of Public Health. The *S*s placed each item into 1 of 11 categories. Standard equal appearing interval instructions were used. The *S*s worked alone to complete the judgment task after receiving appropriate instructions. Item order was newly randomized for each group of cards before an *S* performed the judgment task.

Results. The judgments of three *S*s were discarded due to an apparent reversal of the prescribed method of assigning statements to categories. Median and *Q* values (Edwards, 1957), and means and standard deviations (Diederich et al., 1957), were determined separately for each set of items employing the judgments of the remaining 50 *S*s. Weights of 2, 1, and 0 were used with cutoff points of ±0.8 and ±1.6 in the successive intervals analysis. Plots made separately for each set of items showed near linear relations between successive and equal appearing interval values in all three cases. The plot for hedonic items is shown in Figure 1. Median, mean, *Q*, and standard deviation values for all 98 statements are available (Bruvold, 1967).

Quality and action tendency scales were constructed to provide approximately equal distances between items. Items for the hedonic scale were selected to match those forming the scale commonly used in the sensory evaluation of foods (Schutz, 1965) even though the distance between successive items was not approximately equal. The items comprising

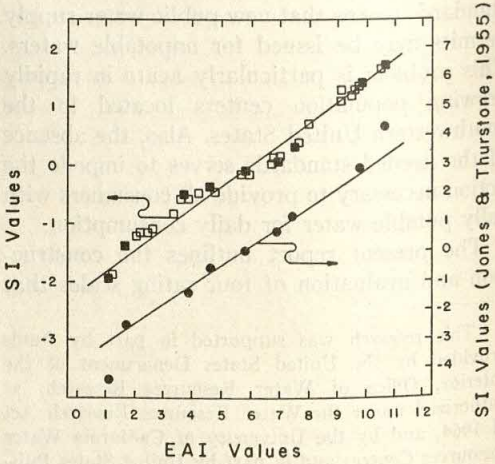


FIG. 1. Plot of equal appearing and successive interval scale values for standard hedonic items. (Circles represent Jones and Thurstone, 1955, data. Filled squares represent items chosen for the water taste hedonic scale. The lines were fitted by eye.)

TABLE 1

ITEMS FOR THE FOUR TASTE SCALES

Hedonic scale	Quality scale
1. I like this water extremely.	1. This water has an excellent taste.
2. I like this water very much.	2. This water has a very good taste.
3. I like this water moderately.	3. This water has a good taste.
4. I like this water slightly.	4. This water has a slightly good taste.
5. I neither like nor dislike this water.	5. This water has a neutral taste.
6. I dislike this water slightly.	6. This water has a slightly bad taste.
7. I dislike this water moderately.	7. This water has a bad taste.
8. I dislike this water very much.	8. This water has a very bad taste.
9. I dislike this water extremely.	9. This water has a horrible taste.
Action Tendency scale	**Combination scale**
1. I would be very happy to accept this water as my everyday drinking water.	1. This water tastes real good. I would be very happy to have it for my everyday drinking water.
2. I would be happy to accept this water as my everyday drinking water.	2. This water tastes good. I would be happy to have it for my everyday drinking water.
3. I am sure that I could accept this water as my everyday drinking water.	3. This water has no special taste at all. I would be happy to have it for my everyday drinking water.
4. I could accept this water as my everyday drinking water.	4. This water seems to have a little taste. I would be satisfied to have it for my everyday drinking water.
5. Maybe I could accept this water as my everyday drinking water.	5. This water has a mild bad taste. I could accept it as my everyday drinking water.
6. I don't think I could accept this water as my everyday drinking water.	6. This water has a fairly bad taste. I think I could accept it as my everyday drinking water.
7. I could not accept this water as my everyday drinking water.	7. This water has a bad taste. I don't think I could accept it as my everyday drinking water.
8. I could never drink this water.	8. This water has a real bad taste. I don't think I would ever drink it.
9. I can't stand this water in my mouth and I could never drink it.	9. This water has a terrible taste. I would never drink it.

each scale are shown in Table 1. Scale and dispersion values are listed in Table 2.

LABORATORY STUDY

Research in a laboratory setting was performed with the four scales in order to obtain an assessment of the reliability of each rating scale and to compare, across scales, the magnitude of mean reliability coefficients. Also the research was designed to yield some information regarding the validity of the scales. In general, judges should rate synthetic

TABLE 2

SCALE AND DISPERSION VALUES FOR SELECTED TASTE SCALE ITEMS

Item no.	Hedonic scale				Quality scale				Action Tendency scale				Combination scale			
	EAI[a]		SI[b]		EAI		SI		EAI		SI		EAI		SI	
	Mdn	*Q*	*M*	*SD*	*Mdn*	*Q*	*M*	*SD*	*Mdn*	*Q*	*M*	*SD*	*Mdn*	*Q*	*M*	*SD*
1	10.66	1.02	1.71	.23	10.67	1.16	2.07	.42	9.96	1.71	1.39	.58	9.54	1.76	1.85	.36
2	9.94	1.49	1.48	.34	9.79	1.31	1.71	.29	9.20	1.86	1.17	.45	8.04	1.67	1.38	.53
3	7.50	1.52	.38	.59	8.45	1.44	1.32	.37	8.07	2.01	.71	.58	6.95	1.11	.81	.50
4	6.98	.60	.04	.23	7.16	.74	.82	.19	7.35	2.10	.50	.52	6.25	2.02	.51	.78
5	5.98	.54	−.20	.05	6.01	.56	.38	.09	5.64	2.02	−.21	.54	5.00	.84	−.02	.38
6	4.87	.66	−.42	.11	4.61	1.11	−.20	.32	4.21	1.63	−.70	.43	4.27	1.83	−.37	.39
7	3.98	1.02	−.56	.17	2.95	1.40	−.79	.32	2.65	1.68	−1.29	.47	3.18	1.86	−.64	.34
8	1.86	1.20	−1.38	.51	2.05	1.22	−1.07	.29	1.27	1.01	−1.87	.47	1.95	1.05	−1.05	.37
9	1.24	1.07	−1.98	.84	1.16	.66	−1.54	.40	1.05	.56	−2.90	1.02	1.33	1.47	−1.61	1.16

[a] Equal appearing intervals.
[b] Successive intervals.

water samples containing a lesser amount of solute more favorably than samples containing a greater amount of the same solute (Bruvold & Pangborn, 1966). Further, it is expected that ratings of equal concentration sodium solutions should vary significantly by anion, with sulfate and bicarbonate solutions receiving the most favorable ratings, chloride intermediate ratings, and carbonate the least favorable ratings (Bruvold & Pangborn, 1966).

Method. Twenty-four individuals served voluntarily as *S*s in the laboratory study. All were employees of the California State Department of Public Health. Twelve *S*s were male and 12 were female. Some *S*s had participated in earlier research which involved rating the taste of water samples. None, however, was asked to serve in the present study on the basis of quality of performance in the earlier work. No *S* was highly experienced in the sensory evaluation of taste, odor, or flavor.

The nine statements comprising a scale appeared, unnumbered, along the left-hand edge of a mimeographed data sheet. The statements were ordered from favorable to unfavorable, starting with the uppermost statement. To the right of each statement there appeared a row of 10 cross marks. The nine rows of 10 cross marks each were also ordered by column to correspond to 10 water samples labeled by code letters. Instructions for the rating procedure appeared at the top of each mimeographed data sheet.

Ten water samples were used in the laboratory work. Double distilled water was used as one sample as was Berkeley tap water. The latter is an odor-free water containing approximately 85 mg/l of total dissolved solids (TDS). Solutions prepared in the laboratory comprised the eight remaining water samples. Reagent grade $NaHCO_3$, Na_2SO_4, NaCl, and Na_2CO_3 were used as solutes, and double distilled water as the solvent, in preparing solutions of 1,000 and 2,000 mg/l. Each solute was represented at each concentration level.

Each *S* took part in nine separate rating sessions. Successive sessions for each *S* were separated by 1 wk. The first session was used only as a practice session. The remaining eight sessions were divided into two blocks of four sessions each. Each scale was used once by each *S* during each block. A list of the 24 possible orders of presentation for the four scales was tabulated. The order in which *S*s used the scales was determined by random matching of *S*s and orders separately for Block 1 and for Block 2. The practice session involved the same scale as Session 1 of Block 1. The average number of interpolated sessions intervening between like scales, over all *S*s, was 3.00 for the hedonic scale, 3.00 for the quality scale, 3.08 for the action tendency scale, and 2.92 for the combination scale.

The actual rating task required *S* to take three separate mouthfuls of each sample, to record his rating on the data sheet by circling the appropriate cross mark, to rinse his mouth thoroughly with Berkeley tap water three times, and to rest for 30 sec. before rating the next sample. The instructions were read to each *S* at the start of the practice session by the experimenter who remained with *S* throughout this initial session to see that all procedures were correctly followed and that the 10 samples were rated in the order specified. The *S*s worked alone in all remaining sessions.

The rating was conducted in a small, quiet, air-conditioned room where the temperature was maintained at 70° ± 2°F. All water samples were served at room temperature in letter-coded 100-ml. beakers filled to the 75-ml. level. Order of sample presentation was randomized for each individual rating session. A different coding system was used for each of the 9 wk. of the study.

TABLE 3

RELIABILITY COEFFICIENTS FOR THE FOUR TASTE SCALES

Reliability coefficient	Frequencies			
	Hedonic scale	Quality scale	Action Tendency scale	Combination scale
.90–.99	1	2	1	2
.80–.89	4	6	4	8
.70–.79	5	8	5	9
.60–.69	4	4	4	4
.50–.59	5	2	5	1
.40–.49	3	1	1	0
.00–.39	2	1	4	0

Results. All analyses performed upon the laboratory results used equal appearing interval scale values as raw data. Standard integer values of one through nine were not employed because of inequality of item spacing especially evident in the hedonic scale.

Pearson product-moment correlation coefficients were computed for all combinations

TABLE 4

AVERAGE INTERCORRELATIONS BETWEEN THE FOUR RATING SCALES

Scale	Hedonic	Quality	Action Tendency	Combination
Hedonic	.63			
Quality	.70	.73		
Action Tendency	.66	.69	.62	
Combination	.70	.72	.71	.77

TABLE 5

MEAN RATINGS FOR WATER SAMPLES USED IN THE LABORATORY STUDY

Solution	Concentration in mg/l	Hedonic scale	Quality scale	Action Tendency scale	Combination scale
Distilled	nil	7.01	7.37	7.60	6.91
Berkeley tap	85	7.59	7.67	8.36	7.33
Na_2SO_4	1,000	6.07	5.79	6.88	6.08
$NaHCO_3$	1,000	5.40	5.15	5.79	5.36
NaC1	1,000	4.85	4.49	5.43	5.33
Na_2CO_3	1,000	3.34	3.11	3.58	3.30
Na_2SO_4	2,000	5.23	4.80	5.90	5.33
$NaHCO_3$	2,000	4.68	4.70	5.23	4.65
NaC1	2,000	3.73	3.61	3.98	3.77
Na_2CO_3	2,000	1.82	1.81	1.95	2.14

of scales and sessions separately for each *S*. Thus, 28 correlation coefficients were obtained per *S*. Each correlation coefficient describes the relation between ratings given identical water samples for a particular pair of rating sessions. Of the 28 correlation coefficients, 4 involve like scales, and therefore these coefficients may be taken as measures of the consistency or reliability of rating by a single *S*. The reliability coefficients for all *S*s and scales are reported in Table 3. Individual reliability coefficients were transformed from r to z values and then subjected to a Treatment (Scales) × Subjects analysis of variance. An F ratio of 4.48 ($df = 3/69$) was obtained for scales. This ratio is significant at the .05 level for a conservative test using 1 and 23 degrees of freedom (Winer, 1962). The average intercorrelations between scales, over all sessions and *S*s, are shown in Table 4.

Mean ratings averaged over sessions and *S*s for each sample and each scale are shown in Table 5. The results of repeated-measures analyses of variance (Winer, 1962), analyses using each *S*'s average rating for each water sample over the two sessions involving like scales as raw data, are shown in Table 6. Sex of *S*, solute concentration, and type of mineral solute formed the three factors in these analyses. Ratings for distilled and Berkeley tap water were not included in the analyses of variance.

CONSUMER SURVEYS

The rating scales developed can be used by *S*s to rate water samples in a laboratory setting, and they can also be used by consumers to rate the taste of water received for daily consumption. The consumer research was designed to provide an assessment of the reliability of the quality rating scale by correlating results from it with attitude scale scores. Further, the surveys were designed to provide some evaluation of the validity of the quality scale by correlating scale ratings with TDS in the waters studied. Theoretical notions concerning consumer attitudes toward water clearly indicate that there should be an imperfect, but substantial, negative correlation between TDS and quality scale ratings (Dillehay et al., 1967).

Method. Four water systems serving communities in the Central California coastal area were selected as testing sites for the quality rating scale. Each

TABLE 6

ANALYSES OF VARIANCE FOR LABORATORY RATINGS

Source	*df*	Hedonic *MS*	Quality *MS*	Action Tendency *MS*	Combination *MS*
Sex (A)	1	0.28	0.40	1.30	2.59
*S*s within groups	22	3.16	1.82	6.61	4.43
Concentrations (B)	1	53.47**	30.13**	64.03**	52.21**
A × B	1	0.16	1.47	1.61	1.05
B × *S*s within groups	22	0.51	1.78	1.15	0.47
Minerals (C)	3	84.30**	76.69**	114.91**	77.85**
A × C	3	0.39	1.59	0.52	0.93
C × *S*s within groups	66	1.30	1.60	2.17	1.30
B × C	3	1.47	2.14	2.82*	1.93*
A × B × C	3	0.79	1.34	1.66	0.21
B × C × *S*s within groups	66	0.95	0.94	1.02	0.62

* $p < .05$.
** $p < .001$.

TABLE 7

DATA FOR THE FOUR COMMUNITIES SURVEYED

Water system	Midrange mineral content in mg/l								Mean ATS	Mean AATS	Mean TSR	No. respondents
	TDS	Ca	Mg	Na	HCO_3	Cl	NO_3	SO_4				
Los Osos	123	5	5	17	32	26	9	3	8.84	8.74	9.80	54*
San Miguel	585	41	42	82	286	73	14	111	7.64	7.37	8.51	55
El Rio	1,001	140	53	80	266	45	12	438	7.28	6.89	7.42	59
Guadalupe	2,078	287	128	122	288	154	14	1,016	5.26	5.16	5.23	51*

* One respondent's TSR missing.

system surveyed was supplied by odor-free water sources as determined by on-site odor detection threshold tests (American Public Health Association, 1965), and no source contained iron or manganese in amounts greater than .05 mg/l. No system chlorinated, and all were supplied by wells whose common mineral content was relatively constant over the system and across the seasons. Chemical analyses and other data for each community water system surveyed are shown in Table 7.

Selection of respondents from each community was accomplished by a careful sampling procedure that involved listing all dwelling units within the geographical area served by the water system under study, the selection of approximately 60 dwelling units by a random sampling procedure modified to distribute selected dwelling units uniformly over the area, and, finally, the random selection of one English-reading individual 15 yr. of age, or older, from all such individuals residing at each selected dwelling unit. Each respondent participating in the survey was first given an interview, then three attitude scales (Dillehay et al., 1967), and finally the quality rating scale. The interview and scales were administered at the respondent's home by a trained interviewer working for the California State Department of Public Health.

Each participating respondent drew a glass of water from his kitchen tap before the taste scale rating procedure was initiated. Having obtained a glass of water, the respondent was given a copy of the quality scale in which the nine items appeared unnumbered and ordered from most to least favorable. The interviewer read each item comprising the scale to the respondent. The respondent was then instructed to take three mouthfuls of water from the glass, to study the scale items carefully between mouthfuls, and to indicate, after taking the third mouthful, which statement most accurately expressed his rating of the taste of the water.

Results. In El Rio, 100% of the selected respondents participated in the survey. Analogous percentages were 93, 95, and 95, respectively, for San Miguel, Los Osos, and Guadalupe.

Intercorrelations between quality taste scale ratings (TSR), attitude taste scale scores (ATS), attitude adjective taste scale scores (AATS), and TDS as milligrams per liter are shown in Table 8. Each correlation coefficient is based upon scale responses obtained from all four towns. Correlation coefficients for taste scale ratings involve 217 respondents, while those for attitude scale scores involve 219 respondents. Equal appearing interval scale values were used in all TSR analyses.

An analysis of linear regression (McNemar, 1962) was performed for each set of scale scores and TDS midranges. The results of the analyses of variance are shown in Table 9. A plot of TSR, ATS, and AATS means against

TABLE 8

INTERCORRELATIONS BETWEEN TDS AND CONSUMER SCALE SCORES

	AATS	ATS	TSR
ATS	.93		
TSR	.83	.84	
TDS	−.63	−.67	−.67

TABLE 9

ANALYSES OF VARIANCE FOR CONSUMER SURVEY RESULTS

Source	ATS		AATS		TSR	
	df	*MS*	*df*	*MS*	*df*	*MS*
Linear regression	1	340.88*	1	333.74*	1	573.63*
Deviation from line	2	2.68	2	4.71	2	2.25
Between-array means	3	115.39*	3	114.38*	3	192.70*
Within arrays	215	1.90	215	2.35	213	3.32
Residual	217	1.91	217	2.37	215	3.31

* $p < .001$.

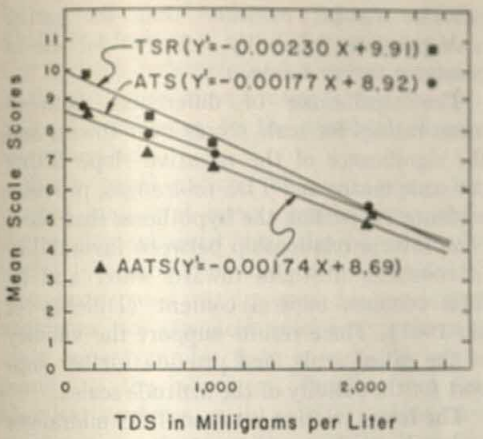


FIG. 2. Lines of best fit for TDS midranges and mean scale scores.

TDS midranges, and the line of best fit for each scale, is pictured in Figure 2.

DISCUSSION

Scale Construction

Edwards (1957) holds that the relation between equal appearing and successive interval scale values for the same set of items will be linear only in the middle range, with successive interval values becoming more extreme near the ends of the continuum. Edwards' prediction was clearly borne out by the relation between successive interval values for general hedonic scale items reported by Jones and Thurstone (1955) and equal appearing interval values for water taste hedonic scale items. This result is pictured in Figure 1. However, none of the plots relating successive and equal appearing interval values for the four sets of items here scaled gave clear support to Edwards' assertion. Instead, the overall linearity obtained between equal appearing and successive interval values shows that the two methods yielded equivalent results for the statements scaled in this research.

Laboratory Study

The reliability coefficients based upon the laboratory work varied substantially across *Ss*. Such a result was not unexpected since *Ss* for this study were volunteers not highly experienced or skilled in sensory evaluation work. Nevertheless, 55 of the 96 individual reliability coefficients were equal to, or greater than, .70. The results indicate that, using water samples similar to those employed in this research, it should be quite easy to obtain a taste panel of *Ss* whose individual reliability coefficients exceed .70 for any of the four scales developed, and that it would be reasonable to seek a panel having individual reliability coefficients of .80, or higher, for the quality or the combination scales.

Schutz (1965) found an action tendency scale to be superior to an hedonic scale in rating food products. In terms of reliability, no such advantage was found for the action tendency scale in the present study. The statistically, and practically, significant differences between mean reliability coefficients were not fully anticipated. The lower reliability of the action tendency scale may be due to the difficulty an *S* encounters in predicting his actual response to a certain mineralized water solely on the basis of laboratory exposure to that water. The lower reliability of the hedonic scale may be due to the rather gross inequality of item spacing along this continuum. Whatever the cause of the findings, the results of the reliability study, the spacing of items, the relative size of the dispersion values, and simplicity of phrasing, all indicate that the quality scale is likely the most useful of the four here investigated. This judgment must be tempered by the purpose of the research in which the scale is to be employed, and also, perhaps, by future findings involving additional assessments of scale quality.

The analyses of variance performed upon *Ss*' average ratings for sodium solutions showed that there was a significant difference between means associated with sample concentration on each of the four scales. This finding is well illustrated by Table 5. Each 1,000 mg/l sample received a more favorable mean rating than the corresponding 2,000 mg/l sample. This result supports the validity of the four scales as it was expected that general taste quality would deteriorate as amount of mineral dissolved in water increased.

The analysis of variance on *Ss*' average ratings also showed that there was a significant *F* ratio associated with solution anions. This result parallels earlier findings with the 14-item combination scale (Bruvold & Pang-

born, 1966). It now seems well established that mean taste scale ratings for equal concentration sodium solutions will clearly rank by anions, with sulfate and bicarbonate ranking highest, chloride intermediate, and carbonate lowest. As pointed out in connection with the earlier work, this rating is identical with the rank order of detection threshold values for sodium salts in water (Bruvold & Pangborn, 1966). The congruence of present findings with earlier ratings and with detection threshold results further supports the validity of the rating scales here described.

Consumer Surveys

The correlations between the quality rating scale and the two attitude scales may be considered as reliability coefficients since both procedures were designed to measure the same attribute. Such a reliability coefficient is most closely related to a coefficient of equivalence obtained by administering equivalent forms of an instrument in close succession. In addition to usual error variance due to chance factors and differences in item content, the present measure of reliability will be affected by differences between the rating and attitude scale methods of obtaining an individual score. Since the obtained correlation coefficients of .83 and .84 are substantial in spite of the additional source of unreliability, it may be concluded that the quality rating scale possesses reliability sufficient for use in survey research of the kind reported in this paper. Further, it should be noted that the correlation between TDS and scale scores was of similar magnitude for both techniques of measurement.

Reliability, in the broader sense of consistency of results, may also be evaluated by comparing the regression of rating and attitude scale scores on TDS. The analyses reported in Figure 2 and Table 9 show that means for all three scales were linearly related to TDS midranges for towns. The rating scale means fit a line of somewhat steeper slope than those fitting attitude scale means; however, such a difference in slopes would not be of much practical importance in recommending a limiting standard for total mineral content in domestic water. The degree of consistency obtained for the regression results provides further evidence that the rating scale possesses reliability sufficient for use in consumer survey research.

The significance of differences between mean ratings for scale scores over towns, and the significance of the negative slope fitting the scale means and TDS midranges, provides evidence supporting the hypothesis that there is an inverse relationship between favorability of consumer attitudes toward water and its total common mineral content (Dillehay et al., 1967). These results support the validity of the rating scale, and provide further support for the validity of the attitude scales.

The linear relation between TDS midranges and rating scale scores for the four water systems studied may appear to be at variance with laboratory work which showed that important differences existed between mean ratings for equal concentration mineral solutions. The apparent difficulty is resolved by noting that, with the exception of bicarbonate for San Miguel and chloride for El Rio, the relations between ionic concentrations and TDS were approximately linear across towns. Nothing in the laboratory work completed to date suggests that a nonlinear relation will be obtained between TDS and mean taste scale ratings when each mineral ion present maintains a relatively constant proportion to TDS. It is also worth noting that the more mineralized natural waters surveyed were relatively low in chlorides and high in sulfates. This fact probably explains the rather moderate mean ratings given El Rio and Guadalupe water. Had these two waters been relatively high in chlorides and low in sulfates, more negative ratings would very probably have been obtained.

CONCLUSIONS

The performance of the quality rating scale in the laboratory and consumer survey research reported here indicates that it may now profitably be employed in future work designed to recommend limiting standards for common mineral content in domestic water (Bruvold et al., 1967). Consumer ratings obtained with the quality scale represent measurement of the attribute of interest by the judgment method (Torgerson, 1958). The previously constructed attitude scales (Dil-

lehay et al., 1967) obtain a measure of the attribute of interest by the response method. Use of both judgment and response methods in the consumer survey research will give additional confidence to the standards finally recommended.

Taste panel ratings obtained using the quality scale and water samples from each community surveyed may also be extremely useful in establishing the desired standards. The taste panel ratings may be considered analogous to chemical analyses performed upon water samples from a particular community in the laboratory. Results from the chemical analyses, such as TDS values, are employed to describe the relation between mineral content in water and potability as operationally defined by consumer attitude or rating scale scores. If it is found that the mean taste panel ratings allow for significant improvement in describing the relation between consumer responses and results from laboratory analyses, taste panel ratings may be incorporated into a regression equation that would be used to predict the potability of mineralized waters not assessed in the consumer survey work. Here the regression equation relating results of laboratory analyses to consumer response comes to embody the desired standard, and, as indicated, inclusion of taste panel data in this equation may considerably increase its accuracy.

The newly developed attitude and rating scales can also be used to assess flavor in water produced by dissolved constituents other than common minerals. In fact, increasing interest in defining and protecting water quality (Bean, 1962) indicates that considerable research will be performed in order to establish standards for taste and odor-producing constituents commonly found in water used for domestic supply. The scales developed make possible an interlocking consumer evaluation-taste panel approach to the assessment of flavor in domestic water. Results from the scales may be used to recommend a limiting standard in terms of an upper limit of acceptability whenever the threshold approach is not feasible (Bruvold et al., 1967). When the threshold approach is both desirable and feasible, assessment of detectable flavor by consumers or taste panel judges will provide fuller knowledge of the effects of the constituent upon water quality. Such additional information would create a more complete data base for recommending the desired standards by indicating the suprathreshold concentration at which a particular constituent in water becomes unacceptable for daily drinking.

REFERENCES

AMERICAN PUBLIC HEALTH ASSOCIATION. *Standard methods for the examination of water and wastewater.* (12th ed.) New York: Author, 1965.

BEAN, E. L. Progress report on water quality criteria. *Journal of the American Water Works Association,* 1962, **54**, 1313–1331.

BRUVOLD, W. H. Four water taste rating scales. Report for Faculty Grant, School of Public Health, University of California, Berkeley, July 1967. (Mimeo)

BRUVOLD, W. H., ONGERTH, H. J., & DILLEHAY, R. C. Consumer attitudes toward mineral taste in domestic water. *Journal of the American Water Works Association,* 1967, **59**, 547–556.

BRUVOLD, W. H., & PANGBORN, R. M. Rated acceptability of mineral taste in water. *Journal of Applied Psychology,* 1966, **50**, 22–32.

DIEDERICH, G. W., MESSICK, S. J., & TUCKER, L. R. A general least squares solution for successive intervals. *Psychometrika,* 1957, **22**, 159–173.

DILLEHAY, R. C., BRUVOLD, W. H., & SIEGEL, J. P. On the assessment of potability. *Journal of Applied Psychology,* 1967, **51**, 89–95.

EDWARDS, A. L. *Techniques of attitude scale construction.* New York: Appleton-Century-Crofts, 1957.

JONES, L. V., & THURSTONE, L. L. The psychophysics of semantics: An experimental investigation. *Journal of Applied Psychology,* 1955, **39**, 31–36.

MCNEMAR, Q. *Psychological statistics.* (3rd ed.) New York: Wiley, 1962.

SCHUTZ, H. G. A food action rating scale for measuring food acceptance. *Journal of Food Science,* 1965, **30**, 365–374.

TORGERSON, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

UNITED STATES PUBLIC HEALTH SERVICE. *Drinking water standards.* Washington, D. C.: Department of Health, Education and Welfare, 1962.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received July 17, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 3, 254–262

HOW TO SUCCEED IN BUSINESS ACCORDING TO BUSINESS STUDENTS AND MANAGERS

BERNARD M. BASS

Management Research Center, University of Rochester

The Organizational Success Questionnaire was developed to see the extent graduate business students and middle managers accept the opinions of the social theorists and the political theorists about what it takes to succeed in large organizations. 6 factors emerged in a 1st analysis but failed to remain sufficiently independent. Consequently, a social approach scale was constructed with a mean internal consistency reliability of .72 and a mean retest reliability of .52. A political scale yielded similar reliabilities. Slight but meaningful relations were found between scale responses and responses to other inventories of orientation and value. Both social and political approaches are needed, fairly or very often, for success according to the respondents (the social somewhat more than the political). But initially students favored less activity of either kind compared to middle managers. Following training the managers remained unchanged but the students became more like the managers.

When your boss asks you to push your subordinates to work harder and your subordinates complain they already have too much to do, what do you do? What do you think the successful executive does? Faced by these and countless other conflicts of interests among those above, below, and next to you in the system, do you concentrate on being open, trusting, and communicative, as probably would be recommended by most socially oriented management experts such as Gibb (1964), Argyris (1962), or Bennis (1964); or, on the contrary, do you become the astute dissembling politician, maintaining your distance, controlling the outflow of information, indulging in minor insincerities, and making superficial compromises, as probably would be recommended by most politically oriented management experts, such as Martin and Sims (1956), Jameson (1945), or Pfiffner (1951)? Gibb argues that before effective interpersonal decision-making procedures can be developed that lead to real commitments, a trusting relationship must be established between the persons. Argyris gives highest priority to "owning up" to one's own feelings and needs. On the other hand, Jameson emphasizes that to be effective in such circumstances, one needs to wear a "mask," while Pfiffner calls for hiding one's own feelings, engaging in benign chicanery, and being a good actor. The present investigation set out to develop a brief attitude scale which would make it possible to see how managers and students felt about the issue, and what effects a particular training program had on these attitudes.

Development of the Organizational Success Questionnaire

As a first step in development of an attitude scale, that is, the Organizational Success Questionnaire (OSQ), 200 undergraduate students enrolled in an elementary psychology section were asked to complete a 16-item questionnaire under the following instructions:

> People differ in what they think it takes to get ahead today in most large organizations. Listed below are ways of behaving that may or may not be important for success. Indicate how frequently you think managers, staff members, or administrators ought to behave as described in order to be successful.

Eight items were composed primarily from an article by Martin and Sims (1956) on the utility of the political approach and eight other statements were drawn from the writings of Gibb (1964) advocating the social approach.

A typical item was:

> Successful persons try to establish in advance mutually satisfactory objectives with others with whom they must work.
>
> 5 — always[1]
> 4 — very often

[1] The frequency scale was based on a magnitude estimation study of 28 adverbs of frequency. Each

3 ___ fairly often
2 ___ sometimes
1 ___ seldom
0 ___ never

Factor Analysis

The 16-item questionnaire was given to 107 graduate business students. A principal components factor analysis with Varimax rotation yielded six factors which accounted for 42% of the variance after 29 iterations (Table 1). All loadings of items on the factors above .40 are italicized in Table 1. The six factors obtained were:

I. Emphasize personal gain.
II. Emphasize organizational gain.
III. Share decision making.
IV. Emphasize candor, openness, trust.
V. Bluff, obscure issues.
VI. Maintain social distance and prerogatives.

For a new sample of 120 graduate business students, factor scores were calculated using the items that correlated most highly with each factor to obtain six factor scores for each student. Table 2 shows the correlations among the factors, as well as correlations with retests scored the same way. It can be seen that Factors I, V, and VI—dealing with personal gain, bluffing, and social distance—clustered together, while Factors II, III, and IV—dealing with organization gain, sharing decisions, and trust—negatively related to the cluster.

The test-retest reliabilities ranging from .35 to .48 indicated that the number of items contributing to each factor score would have to be multiplied many times to achieve reliable assessments of individual attitudes, factor by factor.

Relations of the Factor Scores with Other Assessments

The six factor scores of 113 business students were correlated with a variety

of 71 undergraduate students had been asked to assign a number of his own choosing to "sometimes," then to indicate what number would best fit each of the 28 other adverbs. On the average, in relation to "sometimes," "always" was seen as 2.533 times as frequent; "very often" was 2.093 times as frequent; "fairly often" was 1.683 times as frequent; "seldom" was .425 times as frequent and "never" was .000. Thus, these particular six adverbs of the 28 studied bore an approximate relation to each other of 5:4:3:2:1:0 and were selected as the response alternatives for OSQ items.

TABLE 1

Principal Components Factor Analysis (Decimals Omitted) of 16 OSQ Items Answered by 107 Graduate Business Students

Item	Factor						h^2
	I	II	III	IV	V	VI	
1. Listen only when you ask	−20	*55*	29	32	06	−07	54
2. Level, be candid	34	−14	16	*57*	08	05	50
3. Withhold information	*72*	02	−04	23	−01	−04	57
4. Share decisions	−08	11	*72*	07	−02	−18	58
5. Bluff	21	15	−08	09	*75*	27	72
6. Make allies	*76*	−16	14	−12	01	09	64
7. Commit yourself fully	06	21	−17	*75*	12	−08	67
8. Set mutual objectives	18	18	*40*	13	*−41*	08	42
9. Retard and delay	39	06	−19	18	−30	*46*	52
10. Push people to limits	01	*77*	−05	−14	−13	10	64
11. Maintain distance	−02	02	07	04	10	*84*	73
12. Foster mutual trust	−02	−06	13	*69*	−27	24	62
13. Compromise, but delay	*61*	10	−05	09	*−40*	35	68
14. Hold group discussions	16	−10	*65*	−08	−07	35	59
15. Set clear goals	33	28	03	02	*−65*	21	65
16. Organization comes first	*56*	*51*	08	14	02	−13	62

Note.—All loadings of items on the factors above 40 are italicized.
Factor I = Emphasize personal gain; II = Emphasize organizational gain; III = Share decision making; IV = Emphasize candor, openness, trust; V = Bluff, obscure issues; VI = Maintain social distance and prerogatives.

TABLE 2

INTERCORRELATIONS AMONG THE FACTOR SCORES

Factor	I	II	III	IV	V	VI	Correlation with same factor on retest
I. Personal		−.13	−.13	−.23**	.36**	.50**	.44**
II. Organizational			.17*	.20*	−.24**	−.14*	.35**
III. Share				.00	−.22**	−.12	.39**
IV. Trust					−.22**	−.20	.42**
V. Bluff						.38**	.45**
VI. Distance							.48**

Note.—$N = 120$.
* $p < .05$, $r = .14$ for 118 *df*.
** $p < .01$, $r = .21$ for 118 *df*.

of other assessments. Results are shown in Table 3.

Other assessments. Verbal and numerical aptitude were measured on the Admissions Test for Graduate Schools of Business. Self-, interaction, and task orientation were assessed by the Orientation Inventory (Bass, 1962), a Kuder-type inventory of 27 triads. Values were assessed by the Allport-Vernon Study of Values (Allport, Vernon, & Lindzey, 1960), and budgeting decisions were examined by Exercise Objectives, an exercise in which participants each choose whether or how to spend money on each of five problems: safety, labor relations, management development, product development, and stream clean-up. For each problem, one could choose a "soft" social solution, that is, fix equipment to avoid a possible accident, or a "hard" economic solution, that is, let the machine operator be-

TABLE 3

INTERCORRELATIONS OF OSQ FACTOR SCORES AND ASSESSMENTS OF ABILITY AND VALUES

Assessment	Factor scores					
	I	II	III	IV	V	VI
Verbal aptitude (ATGSB)	.00	.05	−.12	−.01	.10	.06
Numerical aptitude (ATGSB)	.00	−.09	−.24**	.03	.06	−.03
Greater human and social concern in five budgeting decisions	−.20*	.19*	.30**	.25**	−.17*	−.25**
Orientation Inventory						
Self-orientation	.14	−.15*	−.14	−.24**	.05	−.02
Interaction orientation	.06	.08	.08	−.03	.07	.04
Task orientation	−.18*	.04	−.01	.20*	−.10	−.01
Allport-Vernon Study of Values						
Theoretical	−.13	−.14	−.25**	.20*	.02	.01
Economic	.13	−.26**	−.15*	−.07	.08	.19*
Aesthetic	−.11	.08	.10	−.05	.08	−.13
Social	−.05	.10	.15*	.16*	−.09	−.08
Political	.19*	.08	−.04	−.09	−.02	.15
Religious	.04	.15*	.13	.00	.03	−.12

Note.—Factor I = Emphasize personal gain; II = Emphasize organizational gain; III = Share decision making; IV = Emphasize candor, openness, trust; V = Bluff, obscure issues; VI = Maintain social distance and prerogatives.
* $p < .05$, $r = .15$ for 111 *df*.
** $p < .01$, $r = .22$ for 111 *df*.

ware. The total number of "Social" choices was calculated and correlated with OSQ responses.

Several other exercises of the Program of Exercises (Bass, 1967) were also scored against OSQ responses, but failed to yield significant or meaningful patterns of results, possibly because of the low reliabilities of the factor scores.

OSQ and aptitude. Numerical aptitude correlated −.24 with sharing. No other significant relations were observed.

OSQ and orientation. From Table 3 it can be seen that *self-orientation* on the Orientation Inventory—the tendency to seek recognition and personal rewards and to prefer to work alone rather than in groups was negatively related (−.15, −.14, and −.24) with OSQ Factors II, III, and IV, concerned with emphasizing organizational gain, sharing decisions, and fostering trust. On the other hand, *task-orientation*—interest in the challenge of the work to be done—was positively associated (.20) with the OSQ trust score (IV) and also unlike *self-orientation* was negatively related (−.18) with the OSQ factor score for personal gain (I).

OSQ and values. Theoretical value on the Allport-Vernon—appreciation of understanding—was negatively associated (−.25) with sharing of decisions (III) but positively (.20) with trust (IV). Valuing *economic* activity was positively associated with emphasizing social distance (.19) and negatively with emphasizing organizational gain (−.26) and sharing (−.15). Contrary to expectations, strong *social* values were only slightly related to belief in the importance to success of sharing (.15) and trust (.16), while strong *political* values were only slightly related to emphasizing personal gain (.19) and social distance (.15) and not at all to bluffing. Strong religious values correlated .15 with emphasizing organizational gain. While relations here were in the direction expected, it can be seen that the brief OSQ factor scores could in no way substitute for the much more reliable Allport-Vernon Study of Values and would be severely limited to survey and research application.

OSQ and budgeting decisions. Most interesting are the positive correlations of .30 and .25 between the tendency to make "social" budgeting decisions and the OSQ Factors III and IV concerning sharing decisions and fostering trust. Correspondingly, there were three negative correlations of −.20, −.17, and −.25 between these "soft" decision tendencies and belief in the utility of political approaches to organization emphasizing personal gain (I), bluffing (V), and social distance (VI).

Rescoring

The unreliability of the factor scores, their tendency to cluster rather than remain independent, and the patterns of their correlations with other assessments led to the pooling of items from the social factor scores into one *social approach* score and to the pooling of political factor items into one *political approach* score.[2]

Twelve of the original 16 items were clustered into two composites, a social and a political, to form the final scale.

The four items which were eliminated loaded highly on more than one factor or loaded only on Factor II dealing with organizational gain (which could be pursued politically or socially) or loaded highly negatively on a factor.

The final set of 12 items and whether each is scored as social (S) or political (P) are shown in Figure 1.

Scale Reliabilities and Intercorrelations

For a total of 196 cases, test-retest reliabilities were .52 and .51, respectively, for the social and political scores. The correlations between the social and political scales were .49, .63, .33, and .38, respectively, for samples of 200 undergraduates, 24 night business students, 89 day business students, and 83 middle managers. Thus the same respondent who calls for more frequent social approaches to be successful also is likely to call for more frequent political approaches. How much of this may be mere rating bias is unknown.

[2] For two reasons, the alternative of developing six longer, more reliable, and possibly more discriminating factor scales was rejected. First, the social and political dichotomy was in line with the rationale with which the author had begun. Second, the resulting lengthy questionnaire would be much more difficult to administer in the future to large samples.

ORGANIZATIONAL SUCCESS QUESTIONNAIRE

People differ in what they think it takes to get ahead today in most large organizations. Listed below are ways of behaving that may or may not be important for success. Indicate how frequently you think managers, staff members, or administrators ought to behave as described in order to be successful.

Use the following scale to indicate your opinion:

5—They *always* ought to do this.
4—They ought to do this *very often*.
3—They ought to do this *fairly often*.
2—They ought to do this *sometimes*.
1—They *seldom* ought to do this.
0—They *never* ought to do this.

S 1. Level with others; be open, frank and candid in their communications with others.
P 2. Withhold the release of information or time its release for when it will do the most good.
S 3. Share in decision-making with their subordinates whenever possible.
P 4. Act confidently when they are personally unsure about matters or when they lack relevant information.
P 5. Make political alliances with superiors and subordinates to foster and protect mutual interests.
S 6. Completely and openly commit themselves to a position or program.
S 7. Try to establish in advance mutually satisfactory objectives with others with whom they must work.
P 8. Initiate actions which they are personally against but retard and delay carrying out the actions so that the actions are in progress but never completed.
P 9. Maintain social distance; remain aloof, detached, uninvolved with others; always remain the boss when interacting with subordinates.
S 10. Foster mutual trust with others.
P 11. Openly compromise, yet privately divert or delay compromise plans so that their own aims will be pursued despite the stated compromise.
S 12. Arrange for their superiors and their subordinates to meet together to encourage group discussions among others above and below them in the organization with easy participation by all.

FIG. 1. Organizational Success Questionnaire with each item scored as social (S) or political (P).

EFFECTS OF ORGANIZATIONAL EXPERIENCE

Before they entered training programs [3] in organizational psychology, OSQ data were collected from 89 graduate business students, most of whom had had no industrial experience; these data were contrasted with a sample of 24 night students who had had an average of approximately 5 years of experience and a sample of 89 upper-middle managers of large firms who had had about 20 years of experience, on the average.

TABLE 4

HOW TO SUCCEED IN BUSINESS ACCORDING TO GRADUATE BUSINESS STUDENTS AND MIDDLE MANAGERS BEFORE AND AFTER TRAINING

Sample	Mean attitude		
	Pre-training	Post-training	Adjusted post-training
Social items			
83 graduate business students (days)	3.62	3.82	3.90
24 graduate business students (night)	3.56	3.84	3.96
89 middle managers	4.02	4.07	3.93
Political items			
83 graduate business students (day)	3.30	3.26	3.32
24 graduate business students (night)	3.17	3.60	3.73
89 middle managers	3.61	3.69	3.59

Table 4, Column 1, contrasts the mean differences found between the three samples on social approach items, then again on political approach items. All social means were signifi-

[3] The training program focused on completing 10 small group exercises dealing with budgeting, expectations, compensation, life goals, supervision, group dynamics, planning, communication, negotiations, and evaluation. Lectures and T-groups were interwoven in the program. Typically, discussions following each exercise noted that there was no one best way to manage and that a manager had to balance many objectives, be flexible in relating to subordinates, cope with the conflict between individual participation and organizational demands, etc. Neither a complete social nor a complete political point of view was advocated. For more detail, see Bass (1967). The students completed the approximately 40-hour program in 15 weeks; the managers, in 1 week.

cantly higher than the corresponding political means for the same samples.

Table 5, Column 1, contains the two appropriate analyses of the variance of these means. In both cases, the obtained results could not be accounted for by chance. Middle managers felt that to be successful in large organizations required greater frequencies of the social as well as the political approaches described by the items. The mean social score for managers was 4.02, while it was 3.62 and 3.56, respectively, for the student samples. What managers saw as needed *very often* was perceived to be needed between *fairly* and *very often* by the students. Somewhat lower frequencies of political activity were perceived as needed, but again, managers (3.61) felt more of this activity was required for success than did the graduate students (3.30 and 3.17). The correlation of social and political approach means among the three groups probably reflected to some extent the correlation mentioned earlier among the individuals within the groups.

To sum up, prior to training, experienced middle managers called for more frequent use of social and political approaches for success. All three groups favored somewhat greater use of the social approaches compared to the political (3.62 versus 3.30, 3.56 versus 3.30, and 4.02 versus 3.61). All three samples were a bit more in agreement with Gibb, Bennis, and Argyris than with Martin, Sims, Jameson, and Pfiffner. Yet, they subscribed almost as much to the latter men's point of view as to the former.

Specific Differences (before Training)

Table 6 shows how the three samples differed on each of the 12 items studied. Each of the 12 items yielded a standard deviation of .70 to 1.55 with a mean standard deviation of 1.07. Such results signified that almost the whole scale from 0 to 5 was used by respondents. The estimated standard errors of the displayed means ranged from .050 to .076 suggesting that a difference between samples on any of these two means was significant at the 1% level if it was as large as from .128 to .285.

All groups were least disposed toward bluffing, that is, acting confident even when unsure about matters. Probably this political item was the least socially acceptable behavior on the list of 12 items. At the same time, the three samples as a whole were most supportive of the value of fostering mutual trust.

The widest divergence of opinion between the three samples for the various social approach items occurred in the question about

TABLE 5

Analyses of Variance and Covariance of the Social and Political Scores on OSQ of Graduate Business Students and Middle Managers

Source	First analysis			Second analysis		
	df	*SS*	*MS*	Adjusted posttraining *SS*	*MS*	*df*
			Social			
Between samples	2	34.52	103.5	3.28	1.64	2
Error	193	459.31	14.2	1543.31	8.04	192
Total	195	493.83	$F = 7.28^{**}$	1546.60	$F = .20$	194
			Political			
Between samples	2	49.20	197.6	171.00	85.50	2
Error	193	292.07	9.1	2698.36	14.05	192
Total	195	341.27	$F = 21.78^{**}$	2869.40	$F = 6.08^{*}$	194

$^{*}p < .05$.
$^{**}p < .01$.

TABLE 6

SPECIFIC DIFFERENCES BETWEEN GRADUATE BUSINESS STUDENTS AND MIDDLE MANAGERS BEFORE AND AFTER TRAINING

Approved	Before or after training	89 graduate business students (day)	24 graduate business students (night)	89 middle managers	All (N = 196)
Social					
Level, be candid	Before	3.69	3.46	4.20	3.88
	After	3.72	3.85	4.04	3.86
Share decisions	Before	3.19	3.50	4.17	3.64
	After	3.76	4.02	4.23	4.08
Commit oneself fully	Before	3.20	3.13	3.59	3.36
	After	2.98	3.38	3.83	3.38
Set mutual objectives	Before	4.28	3.96	4.35	4.27
	After	4.36	4.29	4.25	4.29
Foster mutual trust	Before	4.58	4.58	4.75	4.65
	After	4.64	4.73	4.86	4.73
Hold group discussions	Before	2.76	2.75	3.06	2.89
	After	3.45	3.28	3.22	3.28
Political					
Withhold information	Before	2.85	2.46	3.25	2.87
	After	3.13	3.63	3.61	3.40
Bluff	Before	1.98	2.21	2.33	2.15
	After	2.06	2.38	2.25	2.18
Make alliances	Before	3.42	3.04	4.01	3.62
	After	3.26	3.38	4.13	3.64
Retard and delay	Before	4.27	4.17	4.37	4.30
	After	3.99	4.13	4.28	4.13
Maintain distance	Before	3.07	3.00	3.31	3.16
	After	3.12	3.83	3.22	3.25
Compromise but divert	Before	4.21	4.17	4.39	4.28
	After	3.97	4.29	4.61	4.29

Note.—5 = always; 4 = very often; 3 = fairly often; 2 = sometimes; 1 = seldom; 0 = never.

the need to share decision making with subordinates whenever possible. The experienced managers saw this as needing to be done *very often* (4.17), the night students with some experience felt less strongly about this (3.50), while the experienced day students called for this action only fairly often (3.17). Again, managers were most in favor of leveling, openness, candor, and full commitment compared to the students.

On political approaches the managers and students were widest apart on the utility of forming political alliances to suit personal ambitions. The managers saw this as needed very often (4.01), the students less so (3.04 and 3.42). The managers also saw the need to withhold information more frequently (3.25) than did the students (2.46 and 2.85).

EFFECTS OF TRAINING IN ORGANIZATIONAL PSYCHOLOGY

Table 4, Column 2, shows the actual mean scores of the three samples following training, and Table 4, Column 3, shows the same

mean scores after the effects have been subtracted from them of the differences among the participants prior to training. Table 5 displays the appropriate covariance analysis of these adjusted posttraining means. From these results we infer that the graduate students, who initially had been significantly ($F = 7.28$) lower than the managers in calling for *social* approaches, increased in this point of view during training while the managers remain unchanged. At the conclusion of training students had become like the managers ($F = .20$) on this dimension. With reference to *political* approaches, both day students as well as the managers remained unchanged in outlook, but as a consequence of training more experienced night students increased in their belief in the efficacy of political activity so much so that their actual final scores on this dimension matched those of the managers and their adjusted mean surpassed that of the managers.

Specific Differences (after Training)

Table 6 shows the responses to each of the 12 OSQ items after as well as before the students and managers completed the course which included the Program of Exercises in Management and Organizational Psychology.

Two multivariate analyses of covariance were completed on these data. In the first, the final six social approach responses were adjusted according to their multiple correlations with the initial social and political responses of the trainees. In the second analysis, the final six political approach responses of the trainees were adjusted according to their multiple correlations with the initial social and political responses of the trainees. In both analyses there remained at the .01 level of confidence significant sources of variance due to whether the trainees were day or night graduate students or middle managers. The respective Fs were 3.85 and 2.54 with 12 and 352 degrees of freedom for each analysis.[4]

Social analysis. The most interesting shifts in stress on social approaches occurred with reference to sharing decisions and holding group discussions. It can be seen that although training raised all samples on how much they thought sharing decisions with subordinates was necessary, the students were affected much more than the managers. The student means rose toward those of what the managers had been. On the other hand, while there was a sharp increase by students in emphasis on holding group discussions, the comparable increase for managers was so much smaller that, after training, students stressed this more than did managers, a reversal of their respective positions before training.

Political analysis. All three samples increased significantly in their valuing the utility of withholding information or timing its release as a consequence of the training effort. Bluffing remained as unpopular as before training. While inexperienced students decreased their support for forming alliances, experienced students and managers increased in their favoring this approach. There was a slight overall decline in all samples in valuing delaying tactics. Among the night graduate students, there was a sharp increase (3.00–3.83) in the opinion that maintaining social distance and aloofness was important for success in the large organization; the other two samples were unchanged. Finally, the middle managers seemed to have increased in their support of open compromise but secret diversion of plans.

Conclusions

A meaningful consistent set of responses emerge when students and managers are asked to indicate what they think it takes to succeed in large organizations. On the whole, they see that the fellow who gets ahead is one who adopts fairly or very often both social and political approaches, the former somewhat more than the latter. He is one who will try to foster trust, share in the decisions with others, and consider the organization's needs. At the same time, to a somewhat lesser extent he is ready when necessary, to maintain his distance and his prerogatives, to act confident, even if he is not sure of himself, and to consider his own personal advantage.

Orientation and values, per se, only slightly relate to these responses, that is, self-oriented respondents favor more political activity, and

[4] For a discussion of the multivariate analysis of covariance see W. W. Cooley and P. R. Lohnes, *Multivariate procedures for the behavioral sciences.* New York: Wiley, 1962.

the attitudes reflected in these responses tend to account for whether socially acceptable rather than economically determined budgeting decisions will be made.

Middle managers tend to see more of both the social and political approaches required of the successful man in the large organization than do graduate business students, but generally the students become more like the managers as a consequence of a training program in organizational psychology while the managers seem less affected by the program.

REFERENCES

ALLPORT, G. W., VERNON, P. E., & LINDZEY, G. *Manual. Study of values.* Boston: Houghton Mifflin, 1960.

ARGYRIS, G. *Interpersonal competence and organizational effectiveness.* Homewood, Ill.: Dorsey Press, 1962.

BASS, B. M. *Orientation Inventory.* Palo Alto: Consulting Psychologists Press, 1962.

BASS, B. M. Combining management training and research. *Training and Development Journal,* 1967, 21(4), 2–7.

BENNIS, W. G. Goals and meta-goals of laboratory training. In W. G. Bennis et al. (Eds.), *Interpersonal dynamics, essays and readings on human interaction.* Homewood, Ill.: Dorsey Press, 1964.

GIBB, J. R. The T-group as a climate for trust formation. In L. P. Bradford, J. R. Gibb, & K. D. Benne, *T-group theory and laboratory methods: Innovation in re-education.* New York: Wiley, 1964.

JAMESON, S. H. Principles of social interaction. *American Sociological Review,* 1945, 10, 6–12.

MARTIN, N. H., & SIMS, J. H. Thinking ahead: Power tactics. *Harvard Business Review,* 1956, 34(6), 25.

PFIFFNER, J. MC. *The supervision of personnel; human relations in the management of men.* New York: Prentice-Hall, 1951.

(Received July 20, 1967)

(Continued from page 239)

Memory Dynamics and Work Motivation: H. Meltzer* and D. Ludwig: Department of Psychology, Washington University, St. Louis, Missouri 63130.

Effect of Lane Closure Signals upon Driver Decision Making and Traffic Flow: Kenneth Perchonok and Paul M. Hurst*: Institute for Research, 257 South Pugh Street, State College, Pennsylvania 16801.

Relationship of Kuder Preference Record V Scores to Adjustment: Implications for Vocational Development Theory: Frank A. Nugent*: Student Counseling Center, Western Washington State College, Bellingham, Washington 98225.

Managerial Attitudes toward Labor Unions in a Southern City: M. Alsikafi,* S. Lee Spray, Walfrid J. Jokinen, and George S. Tracy: Department of Sociology, University of Alabama, P. O. Box 6126, University, Alabama 35486.

Leadership Performance of Nursing Supervisors at Two Organizational Levels: Stanley M. Nealey* and Milton R. Blood: Department of Psychology, University of Illinois, Urbana, Illinois 61801.

Perceived Deficiencies in Individual Need Fulfillment of Career Army Officers: Paul V. Johnson and Robert H. Marcrum*: Office of Military Psychology and Leadership, United States Corps of Cadets, West Point, New York 10996.

Training for Vigilance: Combined Cueing and Knowledge of Results: Earl L. Wiener* and Dennis A. Attwood: Department of Industrial Engineering, University of Miami, Coral Gables, Florida 33124.

---

* Asterisk indicates author for whom the address is supplied.

A Distinguished Series

THE MENTAL MEASUREMENTS YEARBOOKS

The MMY's, the most widely acclaimed reference series in education and psychology, open new vistas in the selection of standardized tests to best serve your specialized needs. In the 6th MMY alone **you get expert assistance in choosing standardized tests by 396 specialists contributing nearly 800 critical reviews of more than 500 tests!**

An indispensable buying guide for increasing numbers of test users, each yearbook supplements rather than supplants earlier volumes. Present MMY volumes include the 3rd ('49, 1062p, $18), 4th ('53, 1188p, $20), 5th ('59, 1321p, $22.50), the 6th ('65, 1752p, $32.50), and *Tests In Print*, ('61, 508p, $7), the master index to the first five MMY's.

Purchasers mentioning this advertisement may obtain the 4 MMY's and *Tests In Print*, at a special price of $90.

Order directly from

THE GRYPHON PRESS 220 Montgomery St., Highland Park, N. J. 08904

A new revision of major significance . . .

APA PUBLICATION MANUAL

1967 Revision

The APA PUBLICATION MANUAL has undergone a complete revision for the first time in 10 years. Although the revised edition prescribes the same basic format for manuscripts as the previous edition, major changes and additions were made in the text in an effort to clarify and simplify the instructions on how manuscripts should be prepared.

The APA PUBLICATION MANUAL is intended primarily as a guide for authors and editors in the preparation of material to be published in APA journals. However, in preparing the original edition it was recognized that the MANUAL could, and should, be written to accommodate a wider range of manuscripts. The revised PUBLICATION MANUAL has retained that original concept.

$1.50

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1200 Seventeenth Street, N. W.
Washington, D. C. 20036

PROCEEDINGS

of the 76th Annual Convention of the American Psychological Association 1968

The 1968 **PROCEEDINGS** will contain the majority of the papers presented at the 1968 Annual Convention and for the first time will contain *summaries of symposia and invited addresses.*

The purpose of **PROCEEDINGS** is to facilitate the dissemination of information regarding current work in major fields of psychology by making research reports available in advance of the Convention and with a minimum publication lag.

The 1968 **PROCEEDINGS**—Volume III of the series which was begun in 1966 with Volume I—will contain approximately 350 reports distributed across the major subject areas of psychology. The reports, together with the summaries of symposia and addresses, will comprise about 700 pages. A publication date of July 1 has been set but you are urged to order your copy of the **PROCEEDINGS** now to avoid a delay in mailing.

PRICE: $12.00 per copy ($6.00 to APA members)

American Psychological Association
1200 17th Street, N. W.
Washington, D. C. 20036

Please send me.....copy(ies) of the 1968 **PROCEEDINGS**

( ) Remittance Enclosed ( ) Bill Me (when issue is mailed)

NAME:.....

ADDRESS:.....

..... Zip Code.....

Copies of the 1966 and 1967 **PROCEEDINGS** are also available at $6.00 and $9.00 respectively. ($3.00 and $4.50 to APA members.)

journal of applied psychology

VOL. 52, NO. 4 AUGUST 1968

JOB SATISFACTION AMONG GRADUATE STUDENTS: INTRINSIC VERSUS EXTRINSIC VARIABLES

EDWARD L. LEVINE AND JOSEPH WEITZ [1]

New York University

A survey of "job" satisfaction was conducted among graduate students in 2 universities. A major source of dissatisfaction at both universities was found to be student voice in influencing department policy. When the responses were factor analyzed, 1st- and 2nd-order factors did not indicate an intrinsic or extrinsic dimension, nor did the findings support a 2-factor theory of "job satisfaction." Furthermore, the alternative to the 2-factor theory, namely that intrinsic variables are more importantly related to measures of overall satisfaction, was similarly not supported. The conclusion is that both the 2-factor theory and its alternative are oversimplifications. Finally, the survey results of general interest are presented.

As in so many recent studies, much of our analysis takes departure from the two-factor theory of job satisfaction outlined by Herzberg, Mausner, and Snyderman (1959). The results of tests of the theory, which was far from exact to begin with (Ewen, 1964), have been mixed, for and against. Burke (1966) offers an objective and fairly thorough summary of much of the relevant literature, while Herzberg (1966), in his more recent book, offers studies supporting the theory and criticizes, sometimes effectively, the research methodology employed by the theory's opponents.

One of the main objections to the two-factor theory is that it is method bound. Most studies which support the theory utilize the same open-ended "critical incidents" approach employed in the original study (cf. Herzberg, 1966, and the review by Hulin, 1966b). On the other hand, evidence against the two-factor theory has come from several sources—studies of dimensionality, including *Q* analyses (Dunnette, Campbell, & Hakel, 1967), factor analysis (Friedlander, 1963; and, as a more direct test, Graen, 1966b), rotated principal components (Ewen, 1964), Coombs' unfolding technique (Burke, 1966), and studies utilizing the overall satisfaction measures neglected by Herzberg et al. (1959) in their original study. Among the latter, analysis of variance (Friedlander, 1966; Graen, 1966a), multiple regression (Halpern, 1966), and a series of hypothesis tests by Ewen, Smith, Hulin, and Locke (1966) have been employed in evaluation of the two-factor theory.

By way of results, the dimensionality studies have not succeeded in establishing any clear-cut intrinsic-extrinsic dichotomy. Furthermore, the other group of studies, along with evidence presented by Wernimont (1966), Dunnette et al. (1967), Hinrichs and Mischkind (1967), and Mischkind (1967), have revealed the weaknesses of the two-factor theory and have pointed to an alternative which postulates that intrinsic (to the job) variables are more important in producing overall satisfaction, as well as overall dissatisfaction.

Our study, a survey of job satisfaction among graduate students majoring in psychology, proposes to extend the work of these investigators. We intend to do a first- as well as a second-order factor analysis, as an addi-

[1] The authors wish to thank Byron Fiman, Jeffrey Greenhaus, Robert Schoen, Theodore Venet, and Virginia Schein who helped construct the questionnaire and collect the data.

© 1968 by the American Psychological Association, Inc.

tion to Ewen's (1964), Friedlander's (1963), and Graen's (1966b) work, in the supposition that an intrinsic-extrinsic dichotomy may emerge in second-order factors.

Moreover, Halpern's (1966) test of the two-factor theory, in which he found that motivators account for most of the variance in overall satisfaction with a satisfied sample, has been vitiated to an extent by the emergence of the above-mentioned alternative theory. His results support the two-factor theory, as well as an alternative theory. The two-factor theory predicts that for people on the satisfied end of the continuum, "motivators" should be more strongly related to overall satisfaction than "hygienes." On the other hand, for people on the dissatisfied end of the continuum, "hygienes" should explain more variance in overall satisfaction.

The alternative theory, which might be called the two-factor-plus-importance theory, predicts that intrinsic variables should relate more strongly to overall satisfaction than extrinsic variables regardless of the level of overall satisfaction; that is, they are more important.

Our results will bear on the two-factor-plus-importance prediction, since, in an extension of Halpern's (1966) test to determine whether intrinsic variables explain more variance in overall satisfaction, we do not restrict our sample only to respondents who are satisfied overall.

Survey data are rich in analytic possibilities, and we would be remiss in reporting only the results bearing on the controversial issues just considered. One interesting area we plan to explore is that of the moderator variable.

Recently, moderator variables have come into prominence as a way of improving predictions (Ghiselli, 1963; Weitz, 1966), and we will employ traits assumed to covary with sex for this purpose.

Hypothesizing that females are more socially oriented than males, and males more "independent," the predictions are: for females, satisfaction with social life, satisfaction with the opportunity for instructor-student discussion, and satisfaction with their social interaction with fellow students will be more highly correlated with overall satisfaction than for males. For males, satisfaction with the voice they have in influencing departmental policy, satisfaction with their income relative to other graduate students, and satisfaction with the opportunity for independent thought and action will be more highly related to overall satisfaction, than for females.

Finally "raw" survey results considered of interest to the scientific community will be presented. Graduate students, of course, will become the psychological community of the future, and their present attitudes toward their training, their projected careers, their mentors, may play a large role in determining the characteristics of that community. To the extent that our results are generalizable, some insights may be obtained from the present data.

PROCEDURE

The Questionnaire

The Graduate Student Questionnaire was designed in the following manner:

1. Of the 78 items on the questionnaire, 35 covered a number of specific areas of possible satisfaction or dissatisfaction along with overall satisfaction. For the majority of the items, a 7-point scale was used, ranging from very satisfied to very dissatisfied. Of the 35 items, 5 were overall measures, for example, "How satisfied are you with your overall graduate training?"; 7 items were rated intrinsic to the students' work, for example, "How satisfied are you with the intellectual stimulation provided by your graduate training?"; 19 items were rated extrinsic to the students' work, for example, "How satisfied are you with your social life?" The remaining 4 items could not be classified into any of the above categories, for example, "How satisfied are you with your decision to become a psychologist?"

The classification of items into intrinsic, or motivator, variables and extrinsic, or hygiene, variables followed the scheme presented by Herzberg et al. (1959). A consensus of opinion among the authors and our colleagues (see Footnote 1) was required for the assignment of items to each of the categories mentioned. Where consensus could not be reached, or where the item did not lend itself to any meaningful classification, no assignment was made.

2. Five items asked what the respondent's primary career goals were. The item format specified that the respondent should divide the 100% of his work time among the categories "teaching," "research in an academic setting," "research in a non-academic setting," "professional practice," "administration," and the ubiquitous "other."

3. The respondents were also asked to suggest the foundation courses which all students in psychology should take.

4. A number of other items covered a variety of topics such as respondent's satisfaction with his grades, to how many other schools he applied and was accepted, whether his present university was his first choice, and whether, if given another chance, he would still choose his present university.

5. Another block of items listed certain professional skills a psychologist might need in his work, for example, experimental design, statistical analysis, etc. The respondent was asked whether this skill was important or not important to him, and whether he was satisfied or dissatisfied with his training in that skill.

6. There were two open-ended questions concerning what the student liked best and least about his graduate training ("Bouquets" and "Brickbats").

7. The last group of items concerned the following demographic characteristics: major area of study, sex, age, marital status, student status, years pursuing graduate study, form of university financial assistance, if any, distance from place of residence to the graduate school, which university the respondent was attending, and whether or not he was a matriculant only (taking no courses), or a full-time student.

This questionnaire, along with a cover letter which explained the purposes of the questionnaire and gave its sponsorship, was mailed to all psychology graduate students at two universities. Anonymity was guaranteed.

Sample

The sample consisted of graduate students enrolled in the psychology department at two universities, one rather large graduate department, and one relatively small.

While the total response rate for the larger university was 53%, the primary results to be presented here will be on 93 cases, all of whom are full-time students taking courses. This represents 70% of the respondents.

For the small university a 61% return was obtained.

RESULTS [2]

Tot. refers to the total group in University A taking courses, $N = 93$.

MO refers to the group maintaining matriculation only, $N = 40$, at University A.

B refers to the second University, $N = 19$.

Factor Analysis

Responses of Tot. on the 35 items concerning specifics of graduate training and overall satisfaction were subjected to a factor analysis. The method of principal axes was used with an Equamax rotation.[3] The orthogonal factors which emerged were then rotated so that they became oblique by a Promax rotation (Hendrickson & White, 1964), and second-order factors were extracted by Varimax (Kaiser, 1958). Squared multiple correlations were placed in the diagonals, and extraction was halted when the number of loadings of items on a factor above .30 was less than two. That the number of factors extracted in the first order is the correct number is given credence by the finding that the number extracted would have been the same if another criterion independent of this one had been used, that is, the number of eigenvalues of the correlation matrix, with unities in the diagonals, greater than one. As many second-order factors were extracted and rotated successively as first-order factors. An upper-bound estimate of the reliability of these items, as given by the Spearman-Brown formula, using the average item intercorrelation (Guilford, 1954), was .91. To the extent that the assumptions of the formula are not met, the coefficient is overstated. However, Guilford (1954, p. 378) points out that a coefficient above .60 will generally not be more than .05 in error, and reliabilities of this magnitude are not extraordinary for items of this type (Edwards, 1957).

Missing data were handled by ignoring those pairs of observations where one or both observations in a pair were not included. The items containing missing data were primarily concerned with assistantship jobs and could not, therefore, be expected to apply to everyone. However, none of the Ns was lower than 70.

The factors themselves emerged as follows: (Only loadings above .30 are considered in view of the small N. Furthermore, the items loading on each factor will be listed by their questionnaire number, and following each item in parenthesis will be the designations I for intrinsic, E for extrinsic, O for overall, and X for unclassifiable measures, as well as the factor loadings.)

Factor 1. General Satisfaction—measures satisfaction with:

[2] The authors gratefully acknowledge the contribution of John D. Hundleby to the analyses reported. Tom Keenan, Chuan-Yu Chen, and Thanton Porter helped prepare the data for computer processing.

[3] David Saunders has devised this criterion.

65. Graduate student life, everything considered (O, .74);
1. Overall graduate training (O, .71);
40. Preparedness to achieve goals (O, .64);
6. Decision to become a psychologist (X, .59);
33. The degree to which the department lived up to expectations (O, .57);
10. Quality of instruction (E, .56);
32. The training in major area of interest (O, .52);
3. Anticipated future as a psychologist (I, .50);
7. Intellectual stimulation from graduate training (I, .48);
8. The intellectual climate in the department (E, .41).

Factor 2. A complicated factor which loaded in order of magnitude on satisfaction with:

25. Academic performance of fellow students (E, .79);
26. Intelligence of fellow students (E, .74);
9. Prestige of department (E, .59);
8. Intellectual climate in the department (E, .51);
23. Work and study interaction with fellow students (E, .45);
7. Intellectual stimulation from training (I, .45);
33. The degree to which the department lived up to expectations (O, .44);
10. Quality of instruction (E, .42);
32. The training in major area of interest (O, .33);
30. The amount of required work in your courses (E, .32).

This factor has been named Satisfaction with Intellectual Stimulation. However, it also seems to measure the quality of peers, the department, and the instruction, along with intellectual stimulation.

Factor 3. Satisfaction with Assistantship Job—measures satisfaction with:

4. Social relationship with your assistantship supervisor (E, .83);
27. Recognition for work as an assistant (I, .81);
28. Feedback on performance as an assistant (I, .64);
15. Income relative to other graduate students (X, .59);
5. Competence of assistantship supervisor (E, .41);
17. Research space available to you (E, .34).

Factor 4. Satisfaction with Physical Environment and Setting—measures satisfaction with:

16. Study space available (E, .72);
18. The overall physical plant at your university (E, .71);
19. Recreational and cultural facilities provided by your university (E, .58);
14. The library facilities (E, .58);
17. Research space available to you (E, .56).

One item, feedback on your performance as an assistant, loaded .40 on this factor; however, the other item loadings were above .55, so that in this case we thought it justifiable to exclude this item.

Factor 5. Satisfaction with Constraints Placed upon the Student—measures satisfaction with:

29. Emphasis placed on grades (E, .67);
22. Freedom in choosing course work (I, .51);
12. Progress toward the PhD (I, .48);
9. Prestige of the department (E, .45);
21. Opportunity for independent thought and action (I, .41);
13. Voice in setting departmental policy (X, .41);
31. Time limits placed on the student (E, .39);
30. Amount of required work in courses (E, .36);
1. Overall graduate training (O, .35);
17. Research space available (E, .32);
11. Opportunity for instructor-student discussion (X, .30).

The only item that causes difficulty is Item 9, prestige of the department. Perhaps we can resolve the presence of this item on the basis that if department prestige is higher the student will have greater latitude in choosing a position, in publishing research, in making contacts. In sum, a department's prestige can "open doors" for the student.

Factor 6. A complicated factor, loading on satisfaction with:

24. Social interaction with fellow students (E, .70);
23. Work and study interaction with fellow students (E, .62);
2. Social life (E, .57);
3. Future as a psychologist (I, .48);
6. Decision to become a psychologist (X, .47);
12. Progress toward the PhD (I, .43).

For want of a better label this factor will be called Satisfaction with Social Future.

Factor 7. Satisfaction with Intellectual Stimulation and Freedom to Pursue Intellectual Interests—measures satisfaction with:

5. The competence of your assistantship supervisor (E, .59);
21. The opportunity for independent thought and action (I, .54);
7. The intellectual stimulation provided by graduate training (I, .46);
8. The intellectual climate in the department (E, .44);
19. The recreation and cultural facilities provided by your university (E, .42);
32. The training in major area of interest (O, .42);
31. The time limits placed on the student (E, .37);
30. The amount of required work in your courses (E, .30);
22. The amount of freedom in choosing course work (I, .30).

The three second-order factors, which account for virtually all the variance in the first-order factors' correlation matrix, are interesting in how they group those of the first order (see Table 1).

Second-Order Factor 1. This factor groups general satisfaction (Factor 1), intellectual stimulation (Factor 2), and constraints placed upon the student (Factor 5). In terms of intrinsic and extrinsic variables, all the intrinsic items (3, 7, 12, 21, 22) are represented here except the two concerning feedback as an assistant and the recognition received as an assistant, while only 10 of 19 extrinsic items appear (8, 9, 10, 17, 23, 25, 26, 29, 30, 31).

TABLE 1

SECOND-ORDER FACTORS

Factor	1	2	3
1	.562	.254	−.003
2	.455	−.096	.297
3	.033	.485	.130
4	.146	.239	.540
5	.570	.156	.056
6	.203	.373	−.241
7	.081	.356	.084

Second-Order Factor 2. This factor groups satisfaction with assistantship job (Factor 3), satisfaction with social future (Factor 6), satisfaction with intellectual stimulation and freedom to pursue intellectual interests (Factor 7). In terms of extrinsic and intrinsic factors, all of the intrinsic items appear (3, 7, 12, 21, 22, 27, 28), while only 10 of 19 extrinsic items appear (2, 4, 5, 8, 17, 19, 23, 24, 30, 31).

Second-Order Factor 3. This factor isolates, to a large extent, satisfaction with physical environment and setting, although satisfaction with intellectual stimulation is represented and, to a small degree, satisfaction with social future. Only two of the seven intrinsic items appear (considering only First-Order Factors 2 and 4 which load .30 or above). Twelve of 19 extrinsic items are represented (1, 9, 10, 14, 16, 17, 18, 19, 23, 25, 26, 30).

Some variance of general satisfaction is accounted for by Second-Order Factors 1 and 2, where all of the intrinsic items appear, but none at all in Second-Order Factor 3 where only 2 of 7 intrinsic and 12 of 19 extrinsic items appear.

Finally, that First-Order Factors 2 and 6, the social factors, were complicated (cf. Hulin, 1966a) is made plainer by the second-order analysis. Factor 2, satisfaction with intellectual stimulation, has some variance represented on Second-Order Factors 1 and 4. Factor 6, satisfaction with social future, has its variance fairly evenly divided among Second-Order Factors 1, 2, and 4. One can only conjecture that the "social" aspects of these dimensions are represented on the more "extrinsic" Second-Order Factor 4, while the more job-related, self-actualizing, intellectual por-

TABLE 2

INTERCORRELATIONS OF OVERALL SATISFACTION MEASURES

Question	1	32	65	40	33
How satisfied are you with:					
1. overall graduate training?	1.00	.49	.66	.50	.57
32. training in major area?		1.00	.48	.44	.62
65. graduate student life, overall?			1.00	.61	.60
40. Will you be prepared to reach your goals?				1.00	.45
33. Did department meet expectations?					1.00

tions are represented in Second-Order Factors 1 and 2, the more "intrinsic" second-order factors.

Interestingly, Item 20, satisfaction with recreational and cultural facilities in the city in which your school is located, does not have a salient loading on any factor, the only variable for which this is true.

Multiple-Regression Analysis

A Wherry stepwise multiple-regression analysis (Wherry, 1940) was employed to determine the relationships of intrinsic and extrinsic items to overall satisfaction. The five overall measures, whose intercorrelations are presented in Table 2, were employed as dependent variables.

The analysis was done in three stages. First, all seven intrinsic variables were correlated with overall satisfaction, then the 19 extrinsic variables, and finally all 26 extrinsic and intrinsic variables together. The multiple-regression coefficients (see Table 3) are obviously in contrast to what the two-factor-plus-importance hypothesis predicts, since the extrinsic variables tend to predict overall satisfaction better in four out of five cases. The differences in coefficients do not seem especially great, but the differences magnify somewhat when one considers the variance accounted for.

TABLE 3

COMPARISON OF THE RELATIONSHIPS OF INTRINSIC AND EXTRINSIC VARIABLES TO OVERALL SATISFACTION MULTIPLE-REGRESSION COEFFICIENTS

Criterion variable[a]	Intrinsic variables	Extrinsic variables
1	.68	.76 (.71)
32	.62	.73 (.69)
65	.72	.79 (.76)
40	.57	.50 (.47)
33	.67	.85 (.83)

Note.—The coefficients in parentheses are the correlations yielded by considering only the best seven extrinsic variables.
[a] See Table 2 for item stems.

A note of caution is in order, for one might well question this evidence on the basis that we "stacked the deck" against the intrinsic variables, since we reported the results for all, and the best seven, variables from the extrinsic set. However, such an objection cannot be sustained when the regression equations of both intrinsic and extrinsic variables on overall satisfaction are computed. The sum of the contributions of each variable to the variance of the overall measures will then give a truer indication of the relative importance of each set of variables. We will arbitrarily set the number of variables included in this analysis to 10, since most of the systematic variance will be accounted for, and the possible effect of the larger number of extrinsic variables will be offset.

The results, as shown in Table 4, are still contrary to the two-factor-plus-importance theory, for in three of five cases extrinsic variables account for more variance. Such a result also implies that the use of a one-item measure of overall satisfaction in tests of a theory is a far from adequate strategy.

Furthermore, while these results were unexpected, they were not unprecedented. Hinrichs and Mischkind (1967) report that, contrary to the two-factor and the two-factor-plus-importance theory, "hygienes" were reported more often than "motivators" as reasons for both satisfaction and dissatisfaction, by both high- and low-satisfaction people.

Survey Results

1. The predictions concerning the moderating effects of sex are presented in Table 5. Two of the predictions hold up: for males, satisfaction with independent thought and action ($p < .05$) and voice in departmental policy ($p < .07$) are more importantly related to overall satisfaction than for females. However, the reverse of the prediction that satisfaction with instructor-student discussion would yield a higher relationship with overall satisfaction for females was borne out ($p < .01$).

2. "Raw" survey results for variables which have been judged important by content, and by their importance to overall satisfaction, reveal the following for B, MO, and Tot.:

All three groups answer all overall satisfaction items "slightly satisfied," with B somewhat more satisfied on the average.

All respondents were slightly satisfied to satisfied, on the average, with their anticipated future as a psychologist, their decision to become a psychologist, the intellectual stimulation provided by their training, the intellectual climate in their department, their progress toward the PhD, the opportunity for independent thought and action, and the amount of required work in their courses. However, the major, and indeed the only outstanding, source of dissatisfaction for all groups was with their voice in influencing department policy, a result consistent with the recent nationwide student unrest.

3. Statistics, psychometrics, and research design, along with history and systems, were endorsed most often by all groups as desirable basic courses for all students in psychology to take.

4. According to the open-ended question requesting "Bouquets" and "Brickbats," the faculty, its general quality, supervisory and technical competence, and breadth of outlook, was the most important source of satisfaction as well as dissatisfaction in both universities.

DISCUSSION

Within the limits set by the sample size, our questionnaire seems to be a fairly reliable instrument, and quite interesting factorially.

TABLE 4

COMPARISON OF THE PROPORTION OF VARIANCE ACCOUNTED FOR BY INTRINSIC AND EXTRINSIC VARIABLES

Criterion variable*	Intrinsic variables	Extrinsic variables
1	.17	.45
32	.13	.46
65	.45	.22
40	.31	.06
33	.01	.71

* See Table 2 for item stems.

It is gratifying that, for the most part, the factors are clear cut; only in Factor 2 and Factor 6 are the interpretations strained, though we are convinced that they are not mere hodgepodges of variables; so that McNemar's (1951) caveat concerning overinterpretation perhaps does not apply here. Furthermore, the factors correspond quite closely to the coding categories set up for the "Brickbat" and "Bouquet" items: freedom, assistantship job, physical environment, level of overall training, peers, are all clearly represented.

Our results, as they bear on the controversy surrounding the two-factor theory and its alternative, add more evidence that the intrinsic-extrinsic dichotomy is not empirically useful. Our Second-Order Factors 1 and 2 can be at best characterized as more intrinsic than extrinsic, and vice versa for Second-Order Factor 3. The complexity of

TABLE 5

COMPARISON OF THE RELATIONSHIPS BETWEEN VARIABLES 2, 11, 13, 15, 17, 21, 24 AND OVERALL SATISFACTION—VARIABLE 1

χ^2(corrected) for males	Variable in question	χ^2(corrected) for females
.04	Social life	.18
6.89; $\phi = .36$; $p < .01$	Instructor student discussion	.28; $\phi = .09$
3.42; $\phi = .25$; $p < .07$	Voice in department policy	1.12; $\phi = .17$
.40	Income relative to others	.43
5.10; $\phi = .32$; $p < .05$	Independent thought and action	.11; $\phi = .06$
.78	Social interaction with peers	.27

this intrinsic-extrinsic dichotomy is further shown by the fact that First-Order Factor 2, a predominantly extrinsic factor, loads most highly with the Second-Order Factor 1, where the general satisfaction factor appears. It seems that the context within which a variable is embedded must be taken into account in terms of its relevance for the satisfaction of self-actualization, or of more basic needs. For example, quality of instruction, for a graduate student, bears heavily on intellectual stimulation and overall satisfaction, while it may mean little to a sales trainee taking a 1-week course. Advancement is quite the "motivator" for an upwardly mobile business executive, but for certain civil servants, whose promotion may depend more upon time, it may have little, if any, motivational implications.

Furthermore, in terms of effects on overall satisfaction, the hypothesis that intrinsic factors are more important, which must make the damaging assumption of an intrinsic-extrinsic dichotomy, does not hold up. Does it not seem strange to expect that, when more sophisticated means of measuring importance have failed (Ewen, 1965), merely classifying variables as intrinsic or extrinsic should allow one to achieve meaningful results?

Nature is frequently unparsimonious. Although a simple explanation is desirable, it must not be so simple as to mask, rather than mirror, the *explicandum* (Chein, 1963; cf. also Dunnette et al., 1967). What is needed in the area of job satisfaction is a broadsided attack on problems of definition and theory where expectations (Weitz, 1956; Wernimont, 1966), community characteristics (Hulin, 1966a), and other situational parameters, including employee needs (Katzell, Barrett, & Parker, 1961), are taken into account. (For the most sophisticated attempts along these lines see Hinrichs and Mischkind, 1967, and Mischkind, 1967, where a model of job satisfaction based on social learning theory is attempted.) Indeed, is it reasonable to approach the study of job satisfaction in a more simple-minded fashion than the study of modern personnel selection and placement (Dunnette, 1963), or to assume that the dynamics at work should be any less complex?

REFERENCES

BURKE, R. J. Are Herzberg's motivators and hygienes unidimensional? *Journal of Applied Psychology*, 1966, **50**, 317–321.

CHEIN, I. The image of man. *Journal of Social Issues*, 1963, **19**, 1–35.

DUNNETTE, M. D. A modified model for test validation and selection research. *Journal of Applied Psychology*, 1963, **47**, 317–323.

DUNNETTE, M. D., CAMPBELL, J. P., & HAKEL, M. D. Factors contributing to job satisfaction and job dissatisfaction in six occupational groups. *Organizational Behavior and Human Performance*, 1967, **2**, 143–174.

EDWARDS, A. L. *Techniques of attitude scale construction*. New York: Holt, Rinehart & Winston, 1957.

EWEN, R. B. Some determinants of job satisfaction: A study of the generality of Herzberg's theory. *Journal of Applied Psychology*, 1964, **48**, 161–163.

EWEN, R. B. Weighting components of job satisfaction. Unpublished doctoral dissertation, University of Illinois, 1965.

EWEN, R. B., SMITH, P. C., HULIN, C. L., & LOCKE, E. A. An empirical test of the Herzberg two-factor theory. *Journal of Applied Psychology*, 1966, **50**, 544–550.

FRIEDLANDER, F. Underlying sources of job satisfaction. *Journal of Applied Psychology*, 1963, **47**, 246–250.

FRIEDLANDER, F. Motivations to work and organizational performance. *Journal of Applied Psychology*, 1966, **50**, 143–152.

GHISELLI, E. A. Moderating effects and differential reliability and validity. *Journal of Applied Psychology*, 1963, **47**, 81–86.

GRAEN, G. B. Addendum to "An empirical test of the Herzberg two-factor theory." *Journal of Applied Psychology*, 1966, **50**, 551–555. (a)

GRAEN, G. B. Motivator and hygiene dimensions for research and development engineers. *Journal of Applied Psychology*, 1966, **50**, 563–566. (b)

GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.

HALPERN, G. Relative contributions of motivator and hygiene factors to overall job satisfaction. *Journal of Applied Psychology*, 1966, **50**, 198–200.

HENDRICKSON, A. E., & WHITE, P. O. Promax: A quick method for rotation to oblique simple structure. *British Journal of Mathematical and Statistical Psychology*, 1964, **17**, 65–70.

HERZBERG, F. *Work and the nature of man*. New York: World Press, 1966.

HERZBERG, F., MAUSNER, B., & SNYDERMAN, B. B. *The motivation to work*. New York: Wiley, 1959.

HINRICHS, J. R., & MISCHKIND, L. A. Empirical and theoretical limitations of the two-factor hypothesis of job satisfaction. *Journal of Applied Psychology*, 1967, **51**, 191–200.

HULIN, C. L. Effects of community characteristics on measures of job satisfaction. *Journal of Applied Psychology*, 1966, **50**, 185–192. (a)

HULIN, C. L. Review of Herzberg, F., Work and the nature of man. *Personnel Psychology*, 1966, **19**, 434–437. (b)

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187–200.

KATZELL, R. A., BARRETT, R. S., & PARKER, T. C. Job satisfaction, job performance, and situational characteristics. *Journal of Applied Psychology*, 1961, **45**, 65–72.

MCNEMAR, Q. The factors in factoring. *Psychometrika*, 1951, **15**, 353–359.

MISCHKIND, L. A. A two-factor theory of job satisfaction; A multi-method and situational analysis. Unpublished doctoral dissertation, New York University, 1967.

WEITZ, J. Job expectancy and survival. *Journal of Applied Psychology*, 1956, **40**, 245–247.

WEITZ, J. Criteria and transfer of training. *Psychological Reports*, 1966, **19**, 195–210.

WERNIMONT, P. F. Intrinsic and extrinsic factors in job satisfaction. *Journal of Applied Psychology*, 1966, **50**, 41–50.

WHERRY, R. J. The Wherry-Doolittle test selection method. In W. H. Stead & C. L. Shartle et al., *Occupational counseling techniques* (Appendix V). New York: American Book Company, 1940.

(Received August 28, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 4, 272–279

CRITERIA USED BY RESEARCH AND DEVELOPMENT ENGINEERS IN THE SELECTION OF AN INFORMATION SOURCE[1]

PETER G. GERSTBERGER AND THOMAS J. ALLEN

Massachusetts Institute of Technology, Sloan School of Management

The criteria employed by engineers in the selection of various technical information channels in problem-solving endeavors are investigated. The relationships of certain criteria identified in past research to such factors as frequency of channel utilization and the rate at which engineers accept or reject technical information received from specific channels are the focal points of the study. A direct relationship is found between perceived accessibility of information channels and several objective measures of utilization, whereas no definite support is found for the hypothesis that the channels perceived highest in technical quality are those used most frequently.

Before information technologists are able to design effective systems to provide the scientist and engineer with the information necessary for his work, a great deal must be learned about the behavior of the user population. Such research is requisite both to assess needs and to improve predictions of the impact on the total information system of new components or modifications of existing ones.

As a result, a large number of what have come to be called "user studies" (investigations of the information-seeking behavior of scientists and engineers) have been performed over the past 10 years. Such studies endeavor to determine (*a*) the effectiveness of currently used information channels, or (*b*) the criteria governing the selection of information channels. Most such studies, until very recently, have not, however, differentiated these two goals. They have implicitly assumed channel effectiveness to be the single criterion governing channel selection. Operating from this assumption, it was necessary only to observe the users and to assign performance measures to the channels on the sole basis of the extent to which they are used.

Allen (1966a, 1966b), applying an independent measure of effective channel performance, found no relation whatever between this measure and the extent to which each of eight channels were used. These results counter any assumption of information quality as the single criterion upon which channel selection is based. Allen proceeds to demonstrate the by now obvious fact that there exists in such a decision process a second parameter of considerable consequence. In addition to the value of a given channel, there is a certain cost associated with using it. The present study is directed at the measurement of relative costs associated with the use of each of nine information channels by R&D engineers and the relative weights assigned cost and value in this decision process (Gerstberger, 1967).

Cost in this sense is a rather complex concept. It may consist of many dimensions—economic, psychological, and physical. This study measures perceived cost to the user in terms of channel accessibility and ease of use. There is considerable support for the idea of measuring cost on these two dimensions. First, Allen's (1966a) data led him to hypothesize that the frequency with which a channel is used is principally determined by the accessibility of the channel to the engineer seeking information. Second, both Werner (1965) and Rosenberg (1966) conclude that the behavior of the users they studied appears to be influenced by perceptions of the ease with which an information channel might be used.

RESEARCH METHODS

The study was conducted in two divisions of a large electronics firm. A sample of 33 engineers was

[1] The research reported in this paper was supported by a grant from the Office of Science Information Service, National Science Foundation (GN 597).

Massachusetts Institute of Technology
Solution Development Record

Name: _____ Date: _____

Problem: _____

a.

alternatives under consideration	estimate of probability that a given alternative will be employed (please circle a probability estimate for each alternative)										
_____	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
_____	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
_____	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
_____	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0

(Circled Probabilities Should Add to 1.0)

b. If technical information having an impact upon your visualization of the problem or any of its potential solutions was received at any time during the past week, please circle the source(s) of that information on the line below.

Information Source(s): L V C ES TS CR G E D

c. Please try to recall two instances during the past week when you were actively searching for technical information related to the problem under consideration. For each case, please fill in a description of the type of information you sought and indicate the *search order* of *all* information channels consulted even though some or possibly all yielded little or no relevant information. (Place a "1" in the space provided in front of the source you consulted first; place a "2" in front of the source you consulted second; and so forth.):

(1) Type of information sought: _____

_____

Search order: ___L___V___C___ES___TS___CR___G___E___D___Other

(please specify): _____

(2) Type of information sought: _____

_____

Search order: ___L___V___C___ES___TS___CR___G___E___D___Other

(please specify): _____

Did you obtain the information sought in (1): ___Yes ___No in (2): ___Yes ___No

Comments: _____

_____

Thank you!

FIG. 1. *The Solution Development Record.* (This device is used to instrument and record activities during Research and Development projects. Respondents report, on a weekly basis, the probability of acceptance of solution alternatives and the use of information sources. Information sources are abbreviated as follows: L = literature; V = vendors; C = customer; ES = external sources; TS = technical staff; CR = company research; G = group discussion; E = experimentation; D = other division. The information sources are explained to the respondent on the reverse side of this device.)

TABLE 1

INFORMATION CHANNELS

Channel	Description
L = literature	Books, professional, technical, and trade journals and other publicly accessible written material.
V = vendors	Representatives of, or documentation generated by, suppliers or potential suppliers of design components.
C = customer	Representatives of, or documentation generated by, the government agency for which the project is performed.
ES = external sources	Sources outside the laboratory or organization which do not fall into any of the above three categories. These include paid and unpaid consultants and representatives of government agencies other than the customer agency.
TS = technical staff	Engineers and scientists in the laboratory who are not assigned directly to the project under consideration.
CR = company research	Any other project performed previously or simultaneously in the laboratory or organization regardless of its source of funding. This includes any unpublished documentation not publicly available, and summarizing past research and development activities.
G = group discussion	Ideas which are formulated as the result of discussion among the *immediate* project group.
E = experimentation	Ideas which are the result of test or experiment or mathematical simulation with no immediate input of information from any other source.
D = other division	Information obtained from another division of your company.

initially selected; of these, 19 remained fully cooperative over the 15-wk. period of the study.[2] Of the 19 cooperating engineers, 8 are employed in one division, 11 in another.[3] Those who are employed in the first division are further subdivided into those who work in a components laboratory and those who are with a systems laboratory. The two laboratories are separated by about ½ mi.; the second division is 18 mi. away. The mean age of the 19 cooperating engineers is 33 yr. and the range 24–43 yr. Six have MS degrees; the remainder all hold BS degrees in engineering. They have been with their present company an average of 5½ yr.

Data were gathered by means of a modified version of the Solution Development Record (Allen, 1966c) (Figure 1). In addition to the questions about the probability of accepting alternative-solution candidates to the *S*'s problem and the use of specific information channels, each *S* was asked each week to report in more detail on two information searches. Specifically, *S*s were asked to report the type of information sought, the order in which information channels were approached, and an indication of the success of the search.

[2] A comparison of the 19 cooperating engineers with the 14 noncooperative ones shows no reason to suspect that they differ in terms of any background characteristics or (from the evidence available) in terms of their information-gathering behavior.

[3] The two divisions operate independently of one another, and the nature of their business is quite different. One division is in the missile systems business; the other is in the business of developing and producing electronic components. All relations reported in this paper, unless otherwise indicated, hold true for both divisions, when each is considered separately.

The modified Solution Development Record was supplemented by periodic questionnaires in which each *S* was asked to rank nine information channels (Table 1) on the basis of the following four criteria:

1. *Accessibility*—accessibility of the nine information channels (without giving consideration to the reliability or quality of the information expected).

2. *Ease of Use*—ease of use of the nine information channels listed.

3. *Technical Quality*—technical quality or reliability of the information obtainable from each of the nine channels listed (without giving consideration to the accessibility of the channel).

4. *Degree of Experience*—degree of experience the engineer has had with the nine channels during his career.

Each of the four channel-ranking questionnaires was administered at a separate point in time over the course of the study. The minimum separation between questionnaires was 1 wk. Upon completion of the study, each *S* was questioned to determine whether he understood clearly the meaning of each of the four dimensions. In a few cases, in which an *S* exhibited confusion over the meaning of a dimension, this difficulty was corrected by the interviewer, and *S* was allowed to rerank the channels.

The resulting rank order of information channels for each criterion provides subjective estimates of "channel cost" and "channel payoff," or value. These channel rankings can be tested for correlation with rank orders based on the frequency-of-use determined from Solution Development Record data. This technique makes possible the identification of the relative importance of various criteria in information-channel selection. The postproject interview was also used to determine the information sources which had

brought to mind each of the solution alternatives considered by each engineer.

RESULTS

Seven hypotheses are tested in the study. All are concerned with predicting the criteria which determine both the frequency with which an information channel is used and the order in which it is consulted by an engineer requiring information:

1. Those information channels which are considered more readily accessible will be used more frequently than less accessible channels.

2. Those information channels which are considered easier to use will be used more frequently than less easy to use channels.

3. Those information channels which are believed to provide information of higher quality will be used more frequently than those providing lower quality, less reliable information.

4. Those information channels which are considered more readily accessible will be used first in the course of an active information search.

5. Those information channels which are considered easier to use will be used first in the course of an active information search.

6. Those information channels which are believed to provide information of higher quality or reliability will be used first in the course of an active information search.

7. A greater proportion of ideas (measured by the ratio of messages accepted to those received) will be accepted from those information channels which are believed to provide information of higher quality or reliability.

Frequency of Use

Testing the first hypothesis, a strong relation (Kendall Tau = .67, $p < .01$) is found (Table 2) between channel accessibility and frequency of use. This first-order test provides substantial support for the first hypothesis. Turning to the second criterion, ease of use, a somewhat weaker relationship ($\tau = .44$, $p < .06$) is found between the channel rankings. The application of partial correlation analysis (Table 3) to the data clearly reveals the magnitude and direction of the relationship. Accessibility is definitely the dominant cost criterion determining the relative frequency with which information channels are used. When accessibility is held constant, ease of use actually shows a slightly negative relation with frequency of use. Similarly, in the case of perceived technical quality, there is a weak positive relation with frequency of use, but when accessibility is controlled, the relation approaches zero. Apparently, in the minds of the *Ss*, there is some relation between their perceptions of tech-

TABLE 2

COMPARISON OF FREQUENCY OF USE WITH MEDIAN ORDERINGS ASSIGNED TO INFORMATION CHANNELS ON THE BASIS OF THREE CRITERIA

Information channel	Criterion			Frequency of Use
	Accessibility	Ease of use	Technical quality	
Literature	3	3	1	1
Vendors	4	5	8	3
Customer	7	8	9	6
External sources	8	9	6	7
Technical staff	2	1	3	4
Company research	6	6	5	9
Group	1	2	4	2
Experimentation	5	4	2	5
Other division	9	7	7	8

Note.—$N = 19$ R&D engineers. Median rank orders for the nine information channels are derived in the following manner. The channel rankings assigned by the 19 engineers for each criterion considered are summed to yield a total score for each of the information sources. For example, if every engineer had ranked "Literature" as being the most accessible channel, the total score for literature would be 19. The channel having the lowest total score is ranked first, the next lowest second, and so forth.

TABLE 3

RELATION OF THREE SELECTION CRITERIA TO FREQUENCY OF INFORMATION CHANNEL USE

Correlation between frequency of use and	Kendall Tau	Kendall Tau (partialed)	Variable partialed out
Perceived accessibility	.67**	.58	Ease of use
		.64	Technical quality
Perceived ease of use	.44*	−.18	Accessibility
		.36	Technical quality
Perceived technical quality	.28	.03	Accessibility
		.08	Ease of use

* $p < .05$.
** $p < .01$.

TABLE 4

RELATION OF THREE SELECTION CRITERIA TO FREQUENCY OF CHANNEL SELECTION AS A FIRST SOURCE

Correlation between frequency of selection as a first source and	Kendall Tau	Kendall Tau (partialed)	Variable partialed out
Perceived accessibility	.67**	.57	Ease of use
		.61	Technical quality
Perceived ease of use	.44*	−.16	Accessibility
		.31	Technical quality
Perceived technical quality	.39	.19	Accessibility
		.21	Ease of use

* $p < .05$.
** $p < .01$.

nical quality and channel accessibility, but it is the accessibility component which almost exclusively determines frequency of use.

First Sources

Each *S* was asked once a week to report two instances during the preceding week in which he actively sought information. In each instance, *S*s reported whether or not the search was "successful" and the order in which they contacted channels. There were 154 searches reported.

Table 4 presents the correlations between frequency of selection as a first source and the three criterion variables. Here again accessibility appears as the dominant criterion upon which selection is based. Engineers turn first to that channel which is most accessible; perceived technical quality influences this decision only to a minor extent.

TABLE 5

MEDIAN RANK ORDERING OF INFORMATION CHANNELS ON THE BASIS OF DEGREE OF EXPERIENCE

Information channel	Ranking assigned
Literature	1
Group	2
Experimentation	3
Technical staff	4
Vendors	5
Company research	6
Customer	7
Other division	8
External sources	9

Note.—N = 19 R&D engineers.

Effect of Experience

One would expect that the degree of experience which an engineer has with a given information channel would influence his perception of both costs and value associated with that channel. He might never use the library, for example, because of a misperception that it contains nothing of value, or that it is really too difficult to find what you are looking for. Given some moderate amount of experience with library use, however, he may come to adjust these perceptions so that they are more in accord with reality. Similarly with interpersonal channels, increased experience should adjust perceived costs. The adjustment can, of course, mean either an increase or decrease in perceived cost, and may be entirely random, increasing for some individuals or channels and decreasing for others. Should the latter situation hold, correlation analysis will fail to detect any systematic trend, and a zero correlation will be found between experience and perceived cost. If, however, experience brings about a systematic shift (most *S*s shifting in the same direction with respect to most channels), correlation

TABLE 6

CORRELATION AND PARTIAL CORRELATION COEFFICIENTS FOR RELATION BETWEEN DEGREE OF EXPERIENCE AND PERCEIVED COSTS OF USING INFORMATION CHANNELS

Correlation between degree of experience and	Kendall Tau	Kendall Tau (partialed)	Variable partialed out
Accessibility	.72**	.37	Ease of use
		.53	Frequency of use
Ease of use	.72**	.37	Accessibility
		.60	Frequency of use
Frequency of use	.61**	.25	Accessibility
		.51	Ease of use
Technical quality	.56*	.43	Accessibility
		.32	Ease of use
		.39	Frequency of use

* $p < .05$.
** $p < .01$.

analysis can detect this phenomenon and will show a positive or negative correlation between experience and cost.

Table 5 shows the nine channels arranged according to the amount of experience the engineers report having had with each of them. Table 6 demonstrates quite clearly that a strong positive relationship holds between the degree of experience an engineer has had with a given channel and both perceived accessibility and ease of use.

Of course, the engineer may simply refer more frequently to those channels which he sees as more accessible or easier to use. In other words, the direction of causality may be from cost to experience. Lower-cost channels are used more, and he thereby accumulates greater experience with them. To test this possibility and determine causal direction, the frequency with which the engineers used the nine channels is controlled and the relation between experience and perceived cost again tested (Table 6). This reduces both of the original relations somewhat, but reasonably strong correlations remain, thus indicating that the degree of experience that an engineer acquires with an information channel does tend to lower his perception of the cost of using that channel.

DISCUSSION

Any assumption that engineers act in accord with a simple instrumental learning model in which they turn most frequently to those information channels which reward them most often should now clearly be laid to rest. Engineers, in selecting among information channels, act in a manner which is intended not to maximize gain, but rather to minimize loss. The loss to be minimized is the cost in terms of effort, either physical or psychological, which must be expended in order to gain access to an information channel.

Their behavior thus appears to follow a "law of least effort" (see, e.g., Zipf, 1949). According to this law, individuals, when choosing among several paths to a goal, will base their decision upon the single criterion of *least average rate of probable work*. In other words, to minimize his average rate of work expenditure over time, ". . . an individual estimates the probable eventualities, and then select(s) a path of least average rate of work through these [Zipf, 1949, p. 6]."

In the selection of information channels, the engineers certainly appear governed by a principle closely related to this. They attempt to minimize effort in terms of the work required to gain access to an information channel. However, they either ignore probable future work (searching through other channels, upon failure to find the needed information through the first channel), or they are unable to estimate it. If, in fact, engineers were to consider future effort in making their decision, one would expect to find a negative correlation between the mean number of channels which must be used to gain the desired information after selecting one channel as the first source. In other words, for each time that a given channel is chosen as a first source, the number (including zero) of channels which had to be used before the information was in hand can be counted. The mean number of additional channels used can be computed for each channel. This mean score serves as an index of the long-term effort (over and above that effort connected with gaining access) associated with each information channel. That correlating this index with the frequency of use of each channel produces a near-zero correlation indicates that engineers do not take this long-term effort into account in selecting channels. So we are left with engineers behaving according to a simplified version of the law of least effort, in which they take only their immediately predictable effort into account and minimize that parameter in making their decision.

The implications of this finding are very important. Improving the quality or performance of a particular information service will not, in and of itself, lead to increased use of the service. More investment in library holdings, for example, will be wasted unless at the same time this material is made more accessible to the user. Engineers will simply not be attracted to the library by improvements in the quality or quantity of the material contained there. The library must, in a sense, come to them.

Those concerned with R&D management whether they are laboratory directors, librarians, or administrators of professional engineering societies have both dimensions of this problem, to varying degrees, within their control. To some extent they can control the quality of the information which is available through certain channels. The laboratory director, for example, can through hiring make available a more competent technical staff to improve the quality of internal consulting. The librarian can improve the extent and quality of the material contained in the library. The professional societies can through careful editing and review improve the quality of their present publications or make new publications or services available to the membership. Improvements in any or all of these directions may well be desirable. But it must be remembered that such improvement, by itself, will not guarantee any improvement in the performance of the organization or of the engineering community. Before the improved information service can lead to increased performance, it must be used! And the only way to increased use is through increased accessibility. Fortunately, however, this parameter can usually be quite readily controlled by management. The laboratory director must not just hire more competent technical staff, he must encourage their use as consultants by rewarding them for this function, and by encouraging his engineers to go to them for information. He must make it known that no one will downgrade a man for seeking help and take steps to reduce the psychological risk and cost in internal consulting. The librarian must make his services more readily available. This may involve some rather expensive additions to personnel (library assistants specializing in particular technical areas, special messenger services, mobile libraries, etc.). But unless this additional investment is made, the money spent on library holdings and facilities will be wasted. The engineering societies, too, can make their services more accessible. Many such improvements, as for example selective dissemination, demand deposit, and the publication of critical annual reviews with extensive reference lists, have been discussed in recent years and many of these hold great promise. But there is a more important step which the societies might take to make their material more accessible. The principal reason for the reluctance of engineers to use their professional literature is that, for the most part, they cannot understand it. Most of the professional engineering literature is too mathematically sophisticated for the average engineer to comprehend. It is therefore inaccessible to him. Now if one wishes to bring engineers into greater contact with their literature, there are two obvious approaches. The first is to operate on the engineer and enhance his mathematical sophistication; equip him so that he can understand the literature as it now exists. The second and perhaps more tractable approach is to bring the literature to the engineer. The professional societies should publish a literature form, whose technical content is reasonably high, but which is understandable by the audience to whom it is directed. This is not an easy task. And it does not necessarily imply that the societies forego all of their present program of publications. What it does mean is that these publications are at least supplemented by a series of more readable works. That this can be done is amply demonstrated by the success of a large number of profit-making, controlled circulation publications, which *engineers do read* (Allen, 1966a). The engineering societies should give up their style of emulating the scientific societies. Their goals are different. There is nothing wrong with publishing nonoriginal contributions in engineering. The goal of engineering publication is to inform, not to stake out claims. Only when the engineering societies come to realize this will they serve their audience properly.

Furthermore, since engineers perceive channels to be more accessible as they gain greater experience in using them, any steps to increase use through improving accessibility will be self-reinforcing. As laboratory managers make their internal consultants more accessible, they will come to be used more and the engineers' perception of their accessibility will be further enhanced. As engineers gain more experience with the more accessible professional literature, their perception of increased accessibility may generalize to the more sophisticated literature and encourage

them to work toward improving their ability to comprehend this literature.

Engineers, it is true, do compensate, by varying the degree of skepticism with which they filter the ideas received, for their lack of attention to quality when selecting an information channel. This is an interesting phenomenon in which technical quality is recognized, but its consideration is delayed until after the channel has been selected. Accepting or rejecting messages is a process which takes place over a considerable period of time. The focus during this time is technical quality (including, of course, cost and hardware availability). The alternatives under consideration are quite rigorously compared in terms of the quality dimension. It is no wonder then that the perceived reliability of the source is taken into account, and that ideas are accepted or rejected, at least in part, on that basis.

From a simple engineering viewpoint, this seems to be an appallingly inefficient way of doing things: using the wrong criterion early, and later correcting for it. This is one more reason why steps must be taken to aid the engineer in overcoming the channel selection problem. Many of the suggestions made earlier in this section should more than pay for themselves in terms of engineering effort conserved.

CONCLUSIONS

Empirical support is found for the model of the information-channel selection process first proposed by Allen (1966a). In this process, channel selection is based almost solely on channel accessibility, the most accessible channels being used first.

The following specific conclusions are derived from the study:

1. Accessibility is the single most important determinant of the overall extent to which an information channel is used.

2. Both accessibility and perceived technical quality influence the choice of first source.

3. Perception of accessibility is influenced by experience. The more experience an engineer has with a channel, the more accessible he perceives it to be.

REFERENCES

ALLEN, T. J. Managing the flow of scientific and technological information. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Sloan School of Management, 1966. (a)

ALLEN, T. J. Performance of information channels in the transfer of technology. *Industrial Management Review*, 1966, 8, 87–98. (b)

ALLEN, T. J. The problem solving process in engineering design. *I.E.E.E. Transactions on Engineering Management*, 1966, 13(2), 72–83. (c)

GERSTBERGER, P. G. An investigation of the criteria used in information channel selection by R&D engineers. Unpublished master's thesis, Massachusetts Institute of Technology, Sloan School of Management, 1967.

ROSENBERG, V. The application of psychometric techniques to determine the attitudes of individuals toward information seeking. Bethlehem, Pa.: Lehigh University Center for the Information Sciences, Report No. 2, 1966.

WERNER, D. J. A study of the information-seeking behavior of medical researchers. Unpublished master's thesis, Northwestern University, 1965.

ZIPF, G. K. *Human behavior and the principle of least effort*. Cambridge, Mass.: Addison-Wesley, 1949.

(Early publication received January 12, 1968)

Journal of Applied Psychology
1968, Vol. 52, No. 4, 280–285

JOB PREFERENCES AND PREFERENCE SHIFTS AS FUNCTIONS OF JOB INFORMATION, FAMILIARITY, AND PRESTIGE LEVEL

RALPH D. NORMAN AND DAVID W. BESSEMER

University of New Mexico

Preference ratings of 10 familiar professional, 10 familiar skilled, 10 unfamiliar professional, and 10 unfamiliar skilled labor jobs, presented as either titles or descriptions, were obtained from 100 male high school students. Professional or familiar jobs were preferred as titles but these effects were absent or reduced in ratings of descriptions of the same jobs. Changes in job preferences in a 2nd rating were greater when Ss were shifted from titles to descriptions, or the reverse, than when the identical type of material was rated, indicating nonequivalence between stimulus materials in study of job attitudes. Implications for personality and interest tests using job titles as items are discussed.

In interest and personality tests such as the Vocational Preference Inventory (Holland, 1965), Minnesota Vocational Interest Inventory (Clark, 1965), and the Strong Vocational Interest Blank (Strong, Campbell, Berdie, & Clark, 1966), the *S* states his occupational preferences. Frequently, however, only job titles are presented to *S* without further information. Quite some time ago, several studies (e.g., Bateman & Remmers, 1939; Nick, 1942; Recktenwald, 1946, 1948) demonstrated the effect of information presented as lectures, discussion, reading, etc., on job-attitude change. Hoppock (1967) argues that providing individuals with such information assumes that attitudes are affected by what they know about vocations. He contends that simple job titles afford only limited cues, and suggests guidelines for providing occupational information, including title, description, qualifications, demands, etc. However, these guidelines are obviously uneconomical when more than a few jobs are to be presented for attitudinal evaluation, and thus have not been used to construct materials for interest tests or job-preference studies.

Information and one of its resultants, familiarity, are not the only variables affecting attitudes toward job titles. Using a few limited titles, Osgood and Stagner (1941) found that occupational stereotypes conform closely to a framework based on prestige, the latter being imputed to titles per se on the basis of brains, being noticed, financial return, excitingness, etc. Moreover, although not immediately relevant to the present study, recent factor-analytic research using titles alone disclosed differences among them on five major factors, which in one study (Gonyea, 1961) failed to resemble conventional interest groupings or job classifications, but which in a second investigation (Gonyea & Lunneborg, 1963) with different methodology resembled a classification proposed by Holland (1959). In a study designed to examine the role played by job descriptions in interaction with titles, Osipow (1962), using the semantic differential, found differences in attitude toward differing job titles (Janitor or Building Superintendent) for the same job on evaluative and potency dimensions which disappeared when descriptions were given. Osipow suggests that perceptions of jobs may be changed with minimal descriptions.

Attitudes toward job titles (vocational stereotypes) have been employed by Holland as the basis of a personality inventory (Holland, 1958) and as a means of developing a theory of personality types based in part on vocational choice. Holland (1966) in fact states:

> The assumption that vocational stereotypes have reliable psychological and sociological meanings is important because most interest inventories rest heavily on its validity. If vocational stereotypes were unreliable and inaccurate, the validity of interest inventories would be seriously reduced. In the case of the present theory, the validity of this assumption lends credence to the first assumption—vocational choice is an expression of personality—and to several subsequent assumptions [p. 5].

The recent work with job titles invites further investigation of variables affecting attitudes toward them, such as familiarity, prestige level, and information provided by job descriptions. Regarding the latter, it was felt that somewhere between the completely restricted cues of titles alone and the detailed guidelines suggested by Hoppock (1967) should lie brief but informative kinds of occupational materials: job descriptions condensed from the Dictionary of Occupational Titles (DOT) (United States Employment Service, 1949).[1] Such descriptions should serve to convey the general "flavor" and essential details of job duties, without being elaborate and time consuming. They should also be particularly useful when job titles are unusual or misleading, and should tend to reduce the influence of prestige level and occupational stereotypes on measures of job attitudes and preferences, as Osipow (1962) clearly demonstrated.

The present study was designed principally to compare titles and descriptions in terms of the influences of job familiarity and prestige level on occupational preference ratings. In addition, results were obtained concerning the stability and equivalence of ratings of these two kinds of materials.

METHOD

Materials

All jobs in the DOT 0 code (professional) and DOT 4 and 5 codes (skilled labor) were scanned, in order to select, on an a priori basis, familiar or unfamiliar jobs in each category. An attempt was made to match at each code level a familiar with one or more quite unfamiliar jobs in the same "job family." For example, Electrician, a skilled familiar job, was paired with Boardman, an unfamiliar job in the same three-digit group. Sixty-eight titles were thus selected, representing a wide range of occupations, and preliminary ratings of familiarity were obtained from small groups of college freshmen (n's of 16 and 17). Inspection of mean ratings caused attrition of 20 titles because no matched job was sharply discriminated on familiarity. The preliminary ratings were also used to determine which matched job offered the greatest familiarity difference for each original title (e.g., Controller was much more unfamiliar than Comptroller as a match for Accountant). The remaining 48 jobs, half professional and half skilled, were then submitted by title to another group of freshmen ($n = 116$) who again rated these jobs on a 7-point scale as to their degree of familiarity with them. By means of the sign test (Siegel, 1956), 40 titles were then selected which were sharply discriminated as to familiarity. These 40 jobs, evenly divided into 10 professional familiar (PF), 10 professional unfamiliar (PU), 10 skilled familiar (SF), and 10 skilled unfamiliar (SU), their DOT codes, and results of the sign test, are presented in Table 1. In all except 1 of the 40 jobs, the match of familiar with unfamiliar occurred exactly for the first three digits of the DOT code, revealing the "job family" closeness of the 20 pairs.

TABLE 1

FAMILIAR AND UNFAMILIAR JOBS, DOT CODES, AND RESULTS OF SIGN TESTS OF FAMILIARITY DIFFERENCES

Familiar	DOT code	Unfamiliar	DOT code	Sign test
Professional jobs				
Accountant	0-01.10	Controller	0-01.70	3.9**
Reporter	0-06.71	Continuity Writer	0-06.23	4.5**
Chemist	0-07.00	Enologist	0-07.10	4.9**
Actor	0-02.11	Elocutionist	0-02.25	4.4**
Medical Doctor	0-26.10	Pathologist	0-26.10	4.1**
Psychologist	0-39.17	Etymologist	0-39.16	4.4**
Criminal Lawyer	0-22.10	County Surrogate	0-22.50	4.5**
Biologist	0-39.31	Ecologist	0-39.33	4.9**
Physicist	0-39.48	Entomologist	0-39.65	4.0**
Civil Engineer	0-16.01	Structural Designer	0-16.01	2.7**
Skilled labor jobs				
Auto Mechanic	5-81.61	Millwright	5-78.10	5.1**
Electrician	4-97.01	Boardman	4-97.11	4.9**
Jeweler	4-71.01	Gemologist	4-71.22	4.4**
Locomotive Engineer	5-41.06	Hostler	5-41.03	4.3**
Gunsmith	5-83.54	Armorer	5-83.54	3.8**
Stone Engraver	4-46.10	Lithograph Artist	4-46.30	2.5*
Welder	4-85.01	Brazer	4-85.31	4.4**
Engraver	4-73.51	Pantographer	4-73.51	4.4**
Stone Mason	5-24.21	Terrazzo Worker	5-24.51	4.4**
Carpenter	5-25.11	Pitwright	5-25.83	5.6**

Note.—DOT = *Dictionary of Occupational Titles* (United States Employment Service, 1949).
* $p < .05$.
** $p < .01$.

In order to have a well-controlled stimulus for job descriptions, each of the longer DOT descriptions of these 40 jobs was condensed to a length of 20–30 words. Mean length was 22.5 words for professional jobs and 23.1 for skilled. In rewriting descriptions, the title of the job was, of course, scrupulously avoided. Thus, the two types of presentation in the main experiment had one which gave the title and no other cues, and one which gave a brief description of the job duties but no title.

Subjects and Design

Four groups of volunteer *S*s rated the 40 occupations on two separate occasions spaced 1 mo. apart.

[1] The 1949 edition of the DOT rather than the latest edition was used in this study in order to assure completely unfamiliar titles, as called for in the design.

The groups differed only in the type of material, that is, titles or descriptions, rated in each session: Group I, Title-Title (TT); Group II, Description-Title (DT); Group III, Title-Description (TD); Group IV, Description-Description (DD). The basic plan was thus a 2 × 2 crossover with respect to type of material and order of presentation, with additional repeated measured factors of familiarity (familiar versus unfamiliar jobs), and prestige level (professional versus skilled jobs) as described above.

A total of 240 *Ss*, all male high school seniors, had been randomly preassigned with 60 in each group. At the first session, 224 *Ss* appeared, with slightly uneven amounts in each group. For administrative reasons, *Ss* were tested as a single group in both rating sessions conducted in the school cafeteria during specially assigned periods. They were aware only of being assigned to one of four groups (not the rationale of the groups) who were given an appropriate booklet in which titles or descriptions, as the case might be, were arranged in random order. The *Ss* were instructed to rate the stimuli on a 5-point Likert-type scale from 1 (like very much) to 5 (dislike very much). In order to stimulate interest prior to rating, they were asked to indicate, in addition to their names and ages, their father's job, own goal job, and college-attendance intent. They were also told that *Es* were interested in their attitudes toward certain jobs, that their cooperation would aid research in this area, and to do their best in rating all stimuli presented to them. The *Ss* were not informed of the type of material rated by other *Ss*. Of the number who participated in the first session, 100 failed to return for the second rating (50 from the "title" condition and 50 from the "description" condition) but no significant differences were observed in the first session data between second session returnees and nonreturnees. An additional 24 *Ss* were randomly eliminated to bring the *n* of each group to equality at 25 *Ss*.

TABLE 2

MEAN PREFERENCE RATINGS OF JOB TITLES AND DESCRIPTIONS

Type of material	Job class	First rating *M*	First rating *SD*	Second rating *M*	Second rating *SD*
		Groups TT & TD		Groups TT & DT	
Titles	PF	3.00	.72	3.03	.76
	PU	3.45	.66	3.50	.72
	SF	3.43	.75	3.43	.74
	SU	3.70	.69	3.76	.69
		Groups DT & DD		Groups TD & DD	
Descriptions	PF	3.16	.59	3.09	.70
	PU	3.50	.53	3.46	.68
	SF	3.24	.62	3.37	.61
	SU	3.42	.58	3.51	.57
Total	PF	3.08	.66	3.06	.73
	PU	3.47	.60	3.48	.70
	SF	3.34	.69	3.40	.68
	SU	3.56	.64	3.64	.63

Note.—Abbreviations: PF = professional familiar; PU = professional unfamiliar; SF = skilled familiar; SU = skilled unfamiliar. TT = Title–Title; TD = Title–Description; DT = Description–Title; DD = Description–Description.

RESULTS

Job Preferences

The mean ratings for each job class and type of material are shown in Table 2. There were no statistically reliable effects attributable to order of presentation, and therefore the data of the groups rating titles and of those rating descriptions, respectively, were combined for the first and second ratings, as the case might be.

The ratings of titles consistently showed professional to be preferred to skilled labor jobs, and familiar to be preferred to unfamiliar jobs. These differences, however, appeared to be considerably reduced in the ratings of descriptions. When titles alone were rated, the difference between means of professional and skilled jobs was .34; the same difference for descriptions was .08. Again, when titles alone were rated, the difference between means of familiar and unfamiliar jobs was .38; the same difference for descriptions was .26. Analysis of variance of the data of the first rating showed significant effects of familiarity ($F = 67.5$, $df = 1/96$, $p < .001$) and prestige level ($F = 7.1$, $df = 1/96$, $p < .01$). Analysis of variance of the second ratings resulted in similar significant effects: familiarity, $F = 111.6$, $df = 1/96$, $p < .001$; prestige level, $F = 10.8$, $df = 1/96$, $p < .01$; but additionally there was a significant interaction between these variables ($F = 10.6$, $df = 1/96$, $p < .01$). It should be noted that no particular meaning can be attached to this interaction since the familiar and unfamiliar jobs were not selected to insure an *equal* difference in familiarity within skilled and professional categories, respectively. While the familiarity effect is somewhat larger for pro-

fessional jobs, there is no reversal of the basic pattern of the main effects.

The only significant effect of type of material found in the first rating involved an interaction with prestige level ($F = 7.1$, $df = 1/96$, $p < .01$). As Table 2 reveals, there was essentially no difference in preference between professional and skilled jobs when these were rated as descriptions, but a considerable difference between them was obtained when they were rated as titles. However, a similar effect was not observed in the data of the second rating ($F = 1.3$, $df = 1/96$, $p > .20$). A significant interaction between type of material and familiarity was obtained in the latter analysis ($F = 5.7$, $df = 1/96$, $p < .05$), apparently reflecting the reduction in the difference between familiar and unfamiliar jobs found for ratings of descriptions as compared to those of titles. This reduction must be attributed largely to the ratings of skilled jobs presented as descriptions, although the three-way interaction was not statistically significant ($F = .23$, $df = 1/96$, $p > .20$).

Preference Shifts

Table 3 shows the mean absolute change in job preferences obtained between the first and second ratings. Since no statistically significant differences were obtained between the two groups rating the same type of material on both occasions, or between the two groups rating different types of material, the data were combined for groups having constant and changed materials, respectively.

The preference shifts were clearly larger when the cue form was changed between the two ratings, as might be expected, but the effect was more pronounced for unfamiliar than for familiar jobs. An analysis of variance confirmed the reliability of these effects, showing significance for change in materials ($F = 15.5$, $df = 1/96$, $p < .001$) and familiarity ($F = 12.5$, $df = 1/96$, $p < .001$), and interaction between these variables ($F = 12.0$, $df = 1/96$, $p < .001$). No other significant effects were obtained in the analysis of the shift scores.

DISCUSSION

Looking at the absolute preference ratings first, Table 2 data show that familiar jobs were obviously preferred to unfamiliar ones, regardless of the occupational prestige level, in both the first and second ratings. This would be comprehensible in the title situation, where the cue was absolutely minimal, that is, one word. However, it is interesting that a similar difference, a sort of "clinging to the familiar," was found in the description situation where there was now far greater verbal elaboration. The only effect of cue form on familiarity was to reduce slightly, but not eliminate, its effect in the second rating. There are two possible reasons for the effect of persistence of the effect of familiarity despite the additional information provided by the description. On the one hand, there would seem to be a tendency for job descriptions of familiar jobs to specify relatively familiar activities with the result that *S* prefers these descriptions to those of the unfamiliar jobs. (This hypothesis is offered despite the fact that in the descriptions the words per se were all within the vocabulary level of high school seniors.) On the other hand, inspection of the job descriptions suggests that those of the familiar jobs, especially at the professional level, involved rather broad, general activities, whereas those of the unfamiliar ones tended to specify relatively narrow, specialized duties. In essence, the latter perhaps provided less opportunity for inclusion of an appealing element to *Ss* such as those of this study.

Regarding prestige level, results showed that job descriptions were effective in eliminating biases in favor of professional jobs found when titles alone were rated. The means

TABLE 3

MEAN PREFERENCE SHIFTS FOR GROUPS RATING CONSTANT AND CHANGED MATERIALS

Job class	Constant Groups TT & DD		Changed Groups TD & DT		Total *M*
	M	*SD*	*M*	*SD*	
PF	.62	.31	.67	.26	.65
PU	.62	.35	.91	.31	.77
SF	.65	.27	.77	.30	.71
SU	.65	.37	.87	.35	.76

Note.—PF = professional familiar; PU = professional unfamiliar; SF = skilled familiar; SU = skilled unfamiliar. TT = Title–Title; DD = Description–Description; TD = Title–Description; DT = Description–Title.

for the first rating were: titles, Professional (P) = 3.23, Skilled (S) = 3.57; descriptions, P = 3.33, S = 3.33. For the second rating the means were: titles, P = 3.27, S = 3.60; descriptions, P = 3.28, S = 3.44. While the prestige level effect was not entirely eliminated by the presentation of descriptions in the second rating as it was in the first, it is the data of the first rating that would appear to have the most relevance to any application of these findings to interest inventories, since such tests are usually taken only once. To summarize here briefly the two main findings, it might be said that job description was more effective in eliminating biases toward prestige level than it was in eradicating the effects of familiarity.

Turning attention now to shift scores, it is obvious that the degree of intersession variation in ratings with the type of material unchanged was comparable across occupational levels and degree of familiarity. Additionally, the finding of equal shifts for the TT and DD groups in the preliminary analyses indicates that ratings of titles and descriptions were equally stable over the interval used here. There was clearly less consistency in ratings with changed cue form, indicating that *S*s generally did not rate titles and descriptions in a similar fashion. This effect was much more pronounced for unfamiliar than for familiar jobs. In fact, it did not seem to appear for the PF jobs (perhaps not quite as obviously so for SF jobs), indicating that *S* recognized these even in description form. The implication of this particular finding is that additional information provided by job descriptions does not compensate for the effects of familiarity very well in terms of mean ratings. It is clear that information influences ratings much more for unfamiliar than for familiar jobs, a situation which would indicate that descriptions are more valid indicators of preferences than titles for unfamiliar jobs. On the other hand, either titles or descriptions are relatively interchangeable as measures for jobs familiar to *S*.

The present study, while admittedly operating with a model of extremes (familiarity *or* unfamiliarity, title *or* description, higher *or* lower prestige level), has some implications for interest inventory item construction, as well as for Holland's (1966) work, which makes extensive use of occupational titles. Regarding interest inventory items, the argument by Kuder (1966) concerning his new Occupational Interest Survey that "occupational titles should be avoided in order to keep obvious vocational significance to a minimum" might carry more weight in view of the findings of this study. As is well known, many of the Kuder items, which involve activities, are in essence exceedingly brief occupational descriptions. Different item formats may thus partially account for reduced correlations between inventories (e.g., Strong versus Kuder) exploring similar vocational interest domains. Fairly recently, King, Norrell, and Powers (1963) studied the relationships between "identical twin" scales (those on the Strong and Kuder bearing the same name) and "fraternal twin" (similar but not identical) scales and found generally that even *r*'s for "identical twin" scales were rather low, the average coefficient being .370.

Concerning Holland's research, since titles alone are apparently sensitive to other variables, such as prestige and familiarity, variations in *S*'s choices (and consequent personality typing) may be influenced by these. For example, in the latest revision of his Vocational Preference Inventory (Holland, 1965) some relatively "unfamiliar" job titles are used, such as Meteorologist, Quality Control Expert, Manufacturer's Representative, Traffic Manager, Inventory Controller, etc. His titles also vary widely over prestige levels. Probably the most prestigious of his six vocational-personality types is the Intellectual, and the least prestigious is the Realistic (Holland, 1966). Many of the job-title preferences of the former closely resemble the PF category in this study; those of the latter resemble the SF category. Whereas the basic validity of his theory of vocational choice is not questioned, it would be valuable to examine the 160 Vocational Preference Inventory titles for significant preference shifts, if any, with job descriptions.

REFERENCES

BATEMAN, R. M., & REMMERS, H. H. Attitudes of high school freshmen toward occupations of their choice before and after studying the occupations

by means of a career book. *Journal of Educational Psychology*, 1939, 30, 657–666.

Clark, K. E. *Minnesota Vocational Interest Inventory*. New York: Psychological Corporation, 1965.

Gonyea, G. C. Dimensions of job perceptions. *Journal of Counseling Psychology*, 1961, 8, 305–312.

Gonyea, G. C., & Lunneborg, C. E. A factor analytic study of perceived occupational similarity. *Journal of Applied Psychology*, 1963, 47, 166–172.

Holland, J. L. A personality inventory employing occupational titles. *Journal of Applied Psychology*, 1958, 42, 336–342.

Holland, J. L. A classification for occupations in terms of personality and intelligence. *American Psychologist*, 1959, 14, 376.

Holland, J. L. *The Vocational Preference Inventory*. Palo Alto: Consulting Psychologists Press, 1965.

Holland, J. L. *The psychology of vocational choice*. Waltham, Mass.: Blaisdell, 1966.

Hoppock, R. *Occupational information*. New York: McGraw-Hill, 1967.

King, P., Norrell, G., & Powers, G. P. Relationships between twin scales on the SVIB and the Kuder. *Journal of Counseling Psychology*, 1963, 10, 395–401.

Kuder, G. F. *Occupational Interest Survey General Manual*. Chicago: Science Research Associates, 1966.

Nick, E. W. High school boys choose vocations. *Occupations*, 1942, 20, 264–269.

Osgood, C. E., & Stagner, R. Analysis of a prestige frame of reference by a gradient technique. *Journal of Applied Psychology*, 1941, 25, 275–290.

Ostrow, S. H. Perceptions of occupations as a function of titles and descriptions. *Journal of Counseling Psychology*, 1962, 9, 106–109.

Recktenwald, L. N. Attitudes toward occupations before and after vocational information. *Occupations*, 1946, 24, 220–223.

Recktenwald, L. N. The appeal of information about occupations. *Journal of Experimental Education*, 1948, 16, 187–191.

Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Strong, E. K., Campbell, D. P., Berdie, R. F., & Clark, K. E. *Strong Vocational Interest Blank for Men*. Stanford: Stanford University Press, 1966.

United States Employment Service. *Dictionary of occupational titles*. Washington, D. C.: United States Government Printing Office, 1949.

(Early publication received March 22, 1968)

Journal of Applied Psychology
1968, Vol. 52, No. 4, 286–289

NEED TYPE AND JOB SATISFACTION AMONG INDUSTRIAL SCIENTISTS[1]

GEORGE B. GRAEN,[2] RENE V. DAWIS, AND DAVID J. WEISS

University of Minnesota

Need-fulfillment theories of job satisfaction generally assume that individuals differ in the outcomes they prefer (need) to obtain from their jobs. Moreover, these theories hypothesize that the relationship between the outcomes received on the job and satisfaction is dependent upon these preferences (needs). The hypothesis this study investigated is that the pattern of preferences for job outcomes moderates the relationship between preference for an outcome and satisfaction with that outcome. Questionnaires were employed to assess preference and satisfaction on 30 different job outcomes. Ss were 113 industrial research scientists working in 1 facility. Preference (need) ratings were subjected to Q cluster analysis. 2 "need" types were identified. It was found that need type moderated the relationship between preference (need) for an outcome and satisfaction with that outcome only on those outcomes most closely related to the company itself. A hypothesis was proposed that these 2 need types reflect the 2 types of research and development personnel that differ primarily in orientation toward the company itself.

The significance of individual differences in motivational variables for the prediction of job satisfaction has been emphasized by several writers (Morse, 1953; Schaffer, 1953; Vroom, 1960, 1964). The rationale commonly associated with such predictions is the need-fulfillment model, which states that job satisfaction is a function of the degree to which needs are met ("fulfilled") by the work environment. According to this model, two individuals with different sets of needs, even though working in the same jobs, with similar working conditions and similar compensation, would not necessarily exhibit the same level of job satisfaction.

In a more recent statement of the need-fulfillment model, the *Theory of Work Adjustment* (Dawis, England, & Lofquist, 1964) hypothesizes that job satisfaction is a function of the correspondence between the needs of the individual and the reinforcer (i.e., need-fulfillment) system of the job. In common with earlier need-fulfillment models, this theory places major emphasis on individual differences in needs, but leaves to empirical research the task of specifying the nature of the relationship between needs and reinforcer systems which is predictive of job satisfaction.

In the present study, a moderator variable is hypothesized as affecting the relationship between need-reinforcer correspondence and job satisfaction. The "need-type hypothesis" proposes that the pattern of needs of the individual is an important moderator variable in the prediction of job satisfaction. More specifically, the hypothesis states that different degrees of accuracy in predicting satisfaction will be found for groups with different need patterns, that is, for different "need types." The need-type hypothesis therefore assumes that there exist different need types, that is, individuals with distinctive need patterns. These need types are seen to differ mainly in the way their needs are interrelated. Thus, need type may be conceptualized as needs organized in complex interaction.

METHOD

To test the need-type hypothesis, *S*s must be working on jobs which have equivalent reinforcer systems. Industrial research scientists working in research and development for one company and at one location may be assumed to meet this requirement. A measure of work-relevant needs also is required for a test of the hypothesis. The Minnesota Importance Questionnaire (Weiss, Dawis, England, & Lofquist, 1964b), an instrument developed to

[1] This study was supported in part by Research Grant RD-1613-G from the Vocational Rehabilitation Administration, United States Department of Health, Education and Welfare, Washington, D. C.

[2] Now at the University of Illinois, Urbana. The senior author wishes to thank Merle E. Ace for critically reading an earlier draft of this manuscript.

measure vocational needs, was employed to meet this requirement.

As the measure of need-reinforcer correspondence, the study employed the subtractive model, which assumes that the more completely the need is "fulfilled," the greater the satisfaction; that is, the strength of the need minus the amount of need fulfillment (reinforcement) equals the degree of satisfaction. Evidence for the moderating effect of need type would be found in differences in the need-satisfaction relationship for different need-type groups or, in terms of analysis of variance, an interaction effect for need-type group and need level on the dependent variable of satisfaction.

The *Ss* were 113 industrial scientists working in research and development for a large company. All worked in the same organizational unit. Median age was 34 yr. and median tenure with the company was 4.3 yr.

The instruments employed were the 1966 revisions of the Minnesota Importance Questionnaire (MIQ) and the Minnesota Satisfaction Questionnaire (MSQ). The MIQ consisted of 117 Likert-type items and was scored on 30 scales. Respondents were asked to rate the importance of various aspects of work to their idea of an ideal job. For each item, the response choices were "not important," "only slightly important," "important," "very important," and "extremely important." The response choices were scored 1 through 5, respectively. Validity evidence for the MIQ as a measure of needs is reported in the *Minnesota Studies in Vocational Rehabilitation* series (Weiss, Dawis, England, & Lofquist, 1964a, 1964b, 1965).

The MSQ consisted of 150 Likert-type items and was scored on 30 scales representing the same reinforcement dimensions utilized for the MIQ.[3] Respondents rated the degree of satisfaction they felt with various aspects of their job. For each item, the response choices were "not satisfied," "only slightly satisfied," "satisfied," "very satisfied," and "extremely satisfied." These were scored 1 through 5, respectively. The MIQ and the MSQ were administered to the *Ss* in the work setting during working hours. To encourage frankness, the identity of the respondents was kept anonymous.

The MIQ scale scores were transformed into *T* scores with a mean of 50 and a standard deviation of 10 to equate scale means and variances. *Q* correlations between each individual and all other individuals were computed for the 113 individuals over the 30 *T*-scored MIQ scales. To identify homogeneous (need type) groups the 113×113 *Q*-correlation matrix was cluster analyzed using Tryon's procedure (Fruchter, 1954).

[3] A table giving reliability coefficients of the MIQ and MSQ scales has been deposited with the American Documentation Institute. Order Document No. 9950 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.25 for photocopies or $1.25 for microfilm and make checks payable to: Chief, Photoduplication Service, Library of Congress.

Once the need-type groups had been identified, 30 two-way analyses of variance for unequal frequencies were performed. The first factor in each ANOVA was need group. The same two categories were used for all ANOVAs. The second factor in the ANOVAs was the MIQ scale corresponding to the same dimension as the dependent variable, the MSQ scale. For all ANOVAs the MIQ scale-score distribution was dichotomized by cutting at the median score. To determine if the moderating effects of need group held in the continuous case, product-moment correlations were calculated between the MIQ scale scores and the corresponding MSQ scale scores for each need group separately. Finally, demographic and MIQ variables were analyzed to determine further the nature of the need-type groups.

RESULTS

Cluster analysis of the 113×113 *Q*-correlation matrix resulted in 12 clusters which included all 113 *Ss*. With the β coefficient set at 1.75, only the first cluster was of sufficient size to warrant use in further analysis. The first cluster contained 32 individuals, the second 12, the third 18, the fourth 13, and the remaining clusters less than 10 individuals each. Therefore, the first cluster of 32 individuals was selected for use as a homogeneous (need-type) group. It was called the "in" cluster. All 81 *Ss* not included in the in cluster represented a heterogeneous need group which was called the "out" cluster.

Of the 30 two-way analyses of variance, 8 yielded significant interactions between need group and need (MIQ), with satisfaction (MSQ) as the dependent variable. Table 1 shows the results of these eight ANOVAs. Only on the Security scale was there a significant need-group mean difference in satisfaction score.

The three largest interaction *F* ratios were found on the dimensions of Working Conditions, Security, and Company Goals. Other scales with significant interactions were Authority, Company Image, Company Policies and Practices, Supervision–Human Relations, and Social Status.

The interaction effect is apparent from a comparison of the means of the four cells. In the case of the in cluster, the high-need group showed higher satisfaction than the low-need group on all eight dimensions. In contrast, for the out cluster, with one exception, the high-need group showed lower satisfaction than the low-need group on these dimensions.

TABLE 1

SATISFACTION MEANS FOR THE EIGHT DIMENSIONS WITH SIGNIFICANT INTERACTIONS

Scale	"In" cluster		"Out" cluster	
	Low need	High need	Low need	High need
Authority	14.36	15.78	15.35	14.30*
Company goals	12.13	14.19	13.29	11.16**
Company image	13.57	16.78	14.86	15.68*
Company policies and practices	13.90	15.83	14.64	13.57*
Security	14.39	17.16	15.02	14.11**
Social status	14.36	15.14	15.57	14.15*
Supervision–human relations	15.47	17.60	16.56	15.05*
Working conditions	15.92	19.26	18.75	17.18**

Note.—$N = 113$.
* Interaction between need group and need level significant at $p \leq .05$.
** Interaction significant at $p \leq .01$.

Correlations between MIQ scores (on a continuous scale) and MSQ scores on the corresponding dimension, for the eight dimensions with significant interactions, are shown in Table 2. Correlations for the total sample ranged from −.20 to .34. These two extreme values were the only correlation coefficients differing significantly from zero. The correlations for the in cluster ranged from .11 to .60 with four of the eight coefficients differing significantly from zero. In contrast, the correlations for the out cluster ranged from −.34 to .23, with four of these being significant. The significant correlations were positive for the in cluster and, with one exception, negative for the out cluster. Fisher's Z transformation was employed to test the differences in correlations between the in cluster and the out cluster. Significant differences were found on seven of the eight dimensions. Only the correlation difference for the Supervision–Human Relations dimension was found to be nonsignificant. The largest difference between correlations was on the Security dimension, which was .60 for the in cluster and −.16 for the out cluster.

An analysis of demographic variables for the two clusters revealed no differences between the groups in age, education, tenure on the job, tenure with the company, time in the profession, job classification, or work-group membership. Similarly, no differences between the means or variances of the in cluster and the out cluster were found on any of the MIQ scales.

DISCUSSION

The hypothesis that need type moderates the prediction of job satisfaction finds support in this study. For the present group of industrial scientists working in research and development for one company, there appears to be at least two need groups for which the relationship between need level and satisfaction on several dimensions is quite different. The improvement in the accuracy of predicting satisfaction contributed by the need-type variable can be seen in Table 2. The correlations between corresponding need and satisfaction dimensions for the entire sample masked the correlations within each of the need groups. For instance, the correlation for the entire sample on the dimension of Security was .03 which masked correlations of .60 for the in cluster and −.16 for the out cluster. Similarly on Working Conditions, the overall correlation was −.01, but for the in cluster and out cluster, respectively, the correlations were .48 and −.20. Thus the moderating effect of the need-type variable in this study is evident.

It is of some interest to inquire into the nature of the two need groups. The in cluster and the out cluster identified in this

TABLE 2

CORRELATION BETWEEN NEED LEVEL AND SATISFACTION ON THE EIGHT DIMENSIONS WITH SIGNIFICANT INTERACTIONS

Scale	Total group ($N = 113$)	Need group		Z_F[a]
		"In" cluster ($N = 32$)	"Out" cluster ($N = 81$)	
Authority	−.01	.23	−.23*	2.15*
Company goals	−.20*	.11	−.34**	2.13*
Company image	.34**	.59**	.23*	2.04*
Company policies and practices	−.06	.43*	−.20	3.05**
Security	.03	.60**	−.16	3.93**
Social status	−.16	.20	−.27*	2.21*
Supervision–human relations	.00	.25	−.09	1.59
Working conditions	−.01	.48**	−.20	3.34**

[a] Z values on Fisher's test for the difference between "in" cluster and "out" cluster correlation coefficients.
* $p \leq .05$.
** $p \leq .01$.

study differed in need-satisfaction relationships on the dimensions of Authority, Company Goals, Company Image, Company Policies and Practices, Supervision–Human Relations, Security, Social Status, and Working Conditions. These are, for the most part, the dimensions most closely associated with satisfaction with the company itself. Among the 30 dimensions included in this study, the only "company" related dimension on which a significant interaction effect was not found was that of Company Prestige. Yet, even on this dimension the correlations between need and satisfaction were in a similar direction, being .15 for the in cluster and −.19 for the out cluster.

These findings suggest that the two need groups might differ in their orientation toward the company. The findings might be interpreted from the viewpoint of the "professionally oriented" and "company oriented" dichotomy for research and development personnel proposed by Schultz (1964). According to Schultz, the professionally oriented view the company merely as a means to achieve their broader professional goals and rewards, whereas the company oriented view the company as the primary source of their achievements and rewards. The in cluster in this study might correspond to the company-oriented type, viewing the company as the major source of need fulfillment. Thus, on the "company" dimensions their satisfaction increases and decreases with need level. In contrast, the out cluster might correspond to the professionally oriented type, or, at least it is not a company-oriented type. Its members apparently do not perceive the company as providing fulfillment of their needs. Thus, the higher their need level, the lower their satisfaction. These data seemingly support the hypothesis of two types of research and development personnel who differ primarily in their orientation toward the company itself. Further study is necessary to test this hypothesis and to determine if this dichotomy can be found in other professional groups employed in industrial firms. Further research also is necessary to determine the generality of the finding reported here that need type operates as a moderator of the relationship between need and satisfaction.

REFERENCES

DAWIS, R. V., ENGLAND, G. W., & LOFQUIST, L. H. A theory of work adjustment. *Minnesota Studies in Vocational Rehabilitation*, 15, University of Minnesota, 1964.

FRUCHTER, B. *Introduction to factor analysis*. New York: Norstrand, 1954.

MORSE, N. D. *Satisfaction in the white-collar job*. Ann Arbor: University of Michigan, Institute for Social Research, Survey Research Center, 1953.

SCHAFFER, R. H. Job satisfaction is related to need satisfaction in work. *Psychological Monographs*, 1953, 67(14, Whole No. 364).

SCHULTZ, D. P. R & D personnel: Two basic types. *Personnel*, 1964, 41, 62–67.

VROOM, V. H. *Some determinants of the effects of participation*. Englewood Cliffs, N. J.: Prentice-Hall, 1960.

VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.

WEISS, D. J., DAWIS, R. V., ENGLAND, G. W., & LOFQUIST, L. H. Construct validation studies of the Minnesota Importance Questionnaire. *Minnesota Studies in Vocational Rehabilitation*, 18, University of Minnesota, 1964. (a)

WEISS, D. J., DAWIS, R. V., ENGLAND, G. W., & LOFQUIST, L. H. The measurement of vocational needs. *Minnesota Studies in Vocational Rehabilitation*, 16, University of Minnesota, 1964. (b)

WEISS, D. J., DAWIS, R. V., ENGLAND, G. W., & LOFQUIST, L. H. An inferential approach to occupational reinforcement. *Minnesota Studies in Vocational Rehabilitation*, 19, University of Minnesota, 1965.

(Received July 17, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 4, 290–295

BRAND AWARENESS:

EFFECTS OF FITTINGNESS, MEANINGFULNESS, AND PRODUCT UTILITY[1]

RABINDRA N. KANUNGO

Dalhousie University, Halifax, Nova Scotia

Comparative influences of fittingness and meaningfulness of the brand name and the utility of the product represented by the brand name on brand awareness were studied. An operational definition of fittingness and a method of selecting fitting brand names were proposed. 24 male and 24 female Ss 1st rated brand names from advertisements and then recalled the brand names. These advertisements included 3 levels of product utility, 2 levels of meaningfulness, and 2 levels of fittingness following a factorial design. Brand awareness was better for (a) highly meaningful and (b) fitting brand names. Meaningfulness influenced brand awareness when brand names were nonfitting, but not when they were fitting. Effects of cultural factors on brand awareness were noticed when brand recall for 3 product-utility categories was analyzed.

Recently, Kanungo and Dutta (1966) reported that brand awareness is better if the brand name is highly meaningful than if it is of low meaningfulness, and that brand awareness of high-utility products is superior to that of low-utility products. In addition, they suggested that the technique of measuring meaningfulness (Kanungo & Panda, 1964a, 1964b; Noble, 1961) of verbal labels could be fruitfully used by the advertisers for the purpose of brand naming. As an extension of this earlier study conducted in India (Kanungo & Dutta, 1966), the present investigation attempts to test the generality of the above findings using subjects (*S*s) belonging to a different cultural milieu.

The present study also examines the effects of the *fittingness* or the *appropriateness* of the brand name for the product on brand awareness. Fittingness of a brand name seems to be an important characteristic that might influence the strength of brand-product association. To the present author's knowledge, no study has been reported on the effects of fittingness variable on brand awareness. The main obstacle to the systematic investigation of the effects of this variable has been the lack of any operational definition and an appropriate measure of fittingness of brand names. Many advertisers often try to select or reject brand names on the basis of their fittingness to their product, but such selections are mostly based upon their intuition or subjective feelings due to the absence of any operational measure. The present study, therefore, tries to offer an operational definition and a measure of the fittingness characteristic of the brand name, so that systematic study of the effect of the variable on brand awareness may be possible.

It is proposed that a brand name may be considered fitting or appropriate for the product if it can readily evoke associations to the product in the minds of the *S*s. Thus, according to the present author, a brand name is fitting for the product if it resembles (in form, sound, or meaning) a word that belongs to one of the most frequent associations to the product. For example, if the product lipstick evokes the word "color" as one of the most frequent associations, then the brand name "Kolory" (apparently a nonsense word, but having resemblance in form and sound to the word "color") may be considered a fitting brand name for lipsticks. However, the same nonsense word will be a nonfitting brand name for the product smoking pipe if the word "color" is only a rare association to the product smoking pipe. This implies that associa-

[1] The study was supported partly by Grant No. X-12-179 from National Research Council of Canada and partly by Grant No. X-85-8 from Dalhousie University Research Development Fund. The author is indebted to Marcia Earhard for many helpful comments.

tion norms for different products may serve as a basis for selecting fitting and nonfitting brand names. Since a fitting brand name would resemble a common associate of the product, one might assume that the initial strength of the brand-product association would be greater in the case of fitting than in the case of nonfitting brand name. Thus with single or repeated exposures brand awareness would be better in the former than in the latter case.

METHOD

Selection of Brand Names

Consideration of the fittingness, the meaningfulness, and the product-utility variable manipulated in the present study influenced the selection of brand names. First, to manipulate the product-utility variable, there was a need to choose products that differed in their utility for the consumers. Thus three different categories of products were selected, namely products used by males, by females, and by both sexes. Four commonly used products from each of these categories were chosen, making a total of 12 products.

Second, in order to choose brand names that differed in their fittingness for the products, the following procedure was employed. Each of the 12 product names was printed on a separate sheet of paper and presented to a group of 19 male and 19 female college undergraduate students. For each product, they were asked to give as many relevant associations as they could within a 30-sec. period. These *Ss* were instructed not to give any existing brand names as their associative responses. All the responses given by the group were found to be meaningful words. The most frequent (occurring 50% or more) and rarest (occurring only once) responses to each product were then determined from the responses given by the group. From these response words, two fitting and two nonfitting brand names for each product were constructed. One fitting and one nonfitting brand name were highly similar to a frequent and a rare response, respectively, and the other fitting and nonfitting brand names only partially resembled in sound or form a frequent and a rare response, respectively. The reason for making half the brand names more similar to the associative responses than the other half was to ensure that the former appear more meaningful than the latter.

Third, to make sure that half of the brand names had higher meaningfulness value than the other half, another 25 undergraduate students were presented with these 48 brand names. These *Ss* were asked to rate each brand name on a 5-point meaningful–meaningless scale. The mean meaningfulness value (m') for each of the brand names was calculated from these ratings. The products, their frequent and rare associates, and the brand names along with their m' are presented in Table 1.

Design and Materials

It will be noticed from Table 1 that for each of the 12 products, there were four brand names: high meaningful-fitting (HM-F), low meaningful-fitting (LM-F), high meaningful-nonfitting (HM-NF), and low meaningful-nonfitting (LM-NF). Using each of these brand names, four different advertisements for each product were prepared. Twelve copies of each of these 48 advertisements were printed on a 14 cm. × 22 cm. art paper. In each printed layout, a brand name following the word "use" appeared at the top, and below which a picture of the product was presented. Below the picture a short phrase containing the product name was printed. For example, in one of the nylon advertisements, at the top of the picture would appear "USE LEGS" and at the bottom of the picture would appear "for your Nylons." Farther down the page, the name and address of the advertiser (a fictitious manufacturing concern) were given.

Finally 48 booklets were compiled. Each booklet contained one advertisement of each of the 12 products in such a manner as to include a HM-F, a LM-F, a HM-NF, and a LM-NF brand name for each of the three product categories: male-use, female-use, and used-by-both products. The sequence of the 12 advertisements in each booklet was randomized to ensure varied order of presentation to *Ss*.

Subjects

Twenty-four adult male and 24 female undergraduate students drawn from two Canadian universities served as *Ss*. Their ages ranged from 18 to 30 yr. The male *Ss* were drawn from St. Mary's University, Halifax, and the female *Ss* were drawn from Mount St. Vincent University, Rockingham. These *Ss* had never participated in any kind of psychological experiment before.

Procedure

Each *S* was given a separate booklet and was asked to rate each brand name, in each advertisement in the booklet, on each of two 6-point scales: meaningfulness and appropriateness. The *S* was given 12 answer sheets, one sheet for each advertisement in the booklet, to record his ratings. The two scales were printed on the answer sheets with the following verbal labels for the 6-points: very high, high, moderate, low, very low, and none. The *S* was instructed to indicate his or her rating by underlining the appropriate verbal label in a scale. The order of presentation of the two scales was counterbalanced for each *S*. To ensure that *Ss* paid proper attention to the brand names in the booklet, each *S* was instructed to write down the brand name he or she was rating at the top of the answer sheet in a blank provided for it. The *Ss* were made to believe that the manufacturer intended to launch an extensive advertising campaign with a view to introducing these new products into the markets. Hence the advertiser wanted to know the reaction in terms of the two scales of various individuals toward these advertisements. The

TABLE 1

PRODUCTS, SOURCE WORDS, AND BRAND NAMES WITH THEIR MEANINGFULNESS (m') VALUE

Product name	High meaningful		Low meaningful	
	Brand name	m'	Brand name	m'
1. Nylons				
F	"Legs" (Leg)	4.36	"Leget" (Leg)	1.28
NF	"Feel" (Feel)	4.12	"Supet" (Superior)	1.20
2. Girdle				
F	"Shapely" (Shape)	4.48	"Tyten" (Tight)	1.24
NF	"Stately" (Stately)	4.28	"Nyten" (Night)	1.28
3. Lipstick				
F	"Colory" (Color)	2.36	"Peenkis" (Pink & Kiss)	1.32
NF	"Shiney" (Shiny)	3.52	"Crimist" (Cream)	1.40
4. Eye shadow				
F	"Blue-Lash" (Blue & Lashes)	3.36	"Loblu" (Blue)	1.36
NF	"Dark-Silk" (Dark & Silk)	3.88	"Lolid" (Lid)	1.28
5. After-shave lotion				
F	"Cooling" (Cool)	4.12	"Smelo" (Smell)	1.40
NF	"Healing" (Heal)	4.32	"Cozic" (Cozy)	1.20
6. Smoking pipe				
F	"Manly" (Man)	4.32	"Tobax" (Tobacco)	1.36
NF	"Clean" (Clean)	4.48	"Nicot" (Nicotine)	1.56
7. Briefcase				
F	"Executive" (Executive)	4.60	"Bizinet" (Business)	1.32
NF	"Important" (Important)	4.32	"Promets" (Promotion)	1.48
8. Shirts (men's)				
F	"Tie-Match" (Tie)	2.48	"Dreso" (Dress)	1.40
NF	"Party-Fit" (Party)	2.52	"Mone" (Money)	1.36
9. Writing pads				
F	"Letters" (Letter)	4.32	"Leterit" (Letter)	1.52
NF	"Economy" (Economy)	4.44	"Thicrap" (Thick)	1.24
10. Pen				
F	"Writer" (Write)	4.36	"Incko" (Ink)	1.32
NF	"Family" (Family)	4.68	"Metto" (Met)	1.40
11. Adhesive tape				
F	"Sticky" (Sticky)	4.24	"Bandes" (Bandaid)	1.48
NF	"Sporty" (Sport)	4.20	"Landes" (Land)	1.40
12. Folders				
F	"Paper-Safe" (Paper)	2.88	"Filex" (File)	1.32
NF	"Camp-Sign" (Camp & Sign)	2.52	"Firex" (Fire)	1.40

Note.—Items 1–4 are female-use products, Items 5–8 are male-use products, and Items 9–12 are products used by both sexes. The bracketed word(s) next to each brand name is the associative response word(s) from which the brand name was derived. F = fitting; NF = nonfitting.

purpose of asking *S*s to rate the brand names was twofold. First, through such ratings, *S*s were given incidental exposure to the advertisements, so that later test of brand awareness would be possible. Second, *S*s' ratings would directly demonstrate if *S*s perceived the meaningfulness and appropriateness characteristic of the brand names in a manner intended by the experimental manipulations.

Each *S* was given approximately 1–1.5 min. to rate each brand name on the two scales. After *S* had finished rating all brand names from the booklet, the answer sheets and the booklet were collected. The *S* was then given a recall sheet on which all the 12 products were listed in alphabetical order and was asked to recall as many brand names as possible and to write them beside the name of the product with which they were associated.

RESULTS

Ratings of Brand Names

The first part of *S*'s task in this experiment was to rate each brand name for its meaningfulness and appropriateness. There were 48 brand names, half of which were HM and the other half LM brand names. Likewise 24 of the 48 brand names were F and the rest NF brand names (see Table 1). The adver-

tisement booklets were distributed in such a manner that each of the 48 brand names was rated by 12 *Ss* (6 male and 6 female *Ss*) on 6-point meaningfulness and appropriateness scales. The 6 points in each scale were assigned ordinal weights of 5 (very high) to 0 (none) for the purpose of deriving the mean meaningfulness and appropriateness value for each brand name. The *S*'s ratings indicated that the 24 HM brand names have significantly higher meaningfulness values (mean $m' = 2.74$) than those of the 24 LM brand names (mean $m' = 1.89$, $t = 3.40$, $p < .01$). The appropriateness ratings also indicate that the F brand names were considered more appropriate (mean appropriateness value = 3.07) than the NF brand names (mean appropriateness value = 1.72, $t = 4.65$, $p < .01$). This suggests that *Ss* did perceive the HM and LM brand names selected by the experimenter as having high and low meaningfulness, respectively. Likewise, they also perceived F and NF brand names, respectively, as more and less appropriate for the products.

Effects of Fittingness and Meaningfulness

In order to determine the effects of fittingness and meaningfulness variables on brand awareness, a $2 \times 2 \times 2$ analysis of variance (Lindquist, 1953, Type VI design) was performed on the recall of brand names. The three classifications were: high and low meaningful brand names, fitting and nonfitting brand names, and male and female *Ss*. The two main effects of meaningfulness and fittingness and their interaction were highly significant (Fs = 23.67, 66.89, and 10.84, respectively). These results are presented in Table 2.

TABLE 2

ANALYSIS OF VARIANCE OF RECALL SCORES

Source of variation	*df*	*MS*	*F*
Between *Ss*	47		
S's sex (S)	1	.15	<1
Error (b)	23	2.65	
Within *Ss*	144		
Meaningfulness (M)	1	17.52	23.67*
Fittingness (F)	1	24.08	66.89*
M × F	1	6.07	10.84*
M × S	1	.96	1.30
F × S	1	.27	<1
M × F × S	1	.35	<1
Error (w)	138	.55	
Error_1 (w)	46	.74	
Error_2 (w)	46	.36	
Error_3 (w)	46	.56	
Total	191		

* $p < .01$.

TABLE 3

MEAN RECALL SCORES FOR FITTING AND MEANINGFUL BRAND NAMES

	Fitting	Nonfitting	*t*
High meaningful	2.81	2.46	2.33*
Low meaningful	2.56	1.50	7.07**
t	1.67	6.40**	

Note.—Means are based on the recall of all 48 *Ss*.
* $p < .05$.
** $p < .01$.

The mean recall of HM–F, LM–F, HM–NF, and LM–NF brand names by all 48 *Ss* presented in Table 3 suggests that brand awareness is better for fitting than for nonfitting brand names and for high meaningful than for low meaningful brand names. Furthermore, these data suggest that meaningfulness variable does not influence brand awareness when the brand names are fitting ones (the difference between HM–F and LM–F recall yields a nonsignificant $t = 1.67$). However, when the brand names are of nonfitting type, the higher the meaningfulness the better is the brand awareness (the difference between HM–NF and LM–NF yields a $t = 6.40$, $p < .01$). This is the reason for the significant interaction between meaningfulness and fittingness. The male and female *Ss* did not differ with respect to their recall of brand names (see Table 2). However, within each group the HM brand names were recalled significantly better than the LM brand names. For male *Ss* the recall means of HM and LM brand names were 5.27 and 3.87, respectively ($t = 6.67$, $p < .01$), and for the female *Ss* the corresponding means were 5.16 and 4.25, respectively ($t = 4.33$, $p < .01$). Similarly for both male and female *Ss* taken separately F brand names were recalled better than NF brand names. For male *Ss* the recall means of F and NF brand names were 5.21 and 3.83,

respectively ($t = 6.57$, $p < .01$), and for female *S*s the corresponding means were 5.33 and 4.08, respectively ($t = 5.95$, $p < .01$).

Effect of Utility of the Product

A 2 × 3 analysis of variance was performed on the recall scores of male and female *S*s for the brand names of the three categories: male-use, female-use, and used-by-both products. Out of the two main effects, the male and female *S*s, and three product categories, only the latter approached significance ($F = 2.74$, $df = 2/92$, $.10 > p > .05$). The mean recall scores of the male *S*s were 3.29, 3.12, and 2.87, and the corresponding recall scores of the female *S*s were 3.17, 3.25, and 3.00 for male-use, female-use, and used-by-both products, respectively. These results indicate that male and female *S*s did not differ in their recall patterns. The brand names for male-use and female-use products were recalled equally well by both the groups. There was a trend in both the groups to recall fewer brand names for the used-by-both products than those for the other two categories.

DISCUSSION

The above results clearly confirm the earlier findings that brand names having higher meaningfulness values are retained better than those having low meaningfulness value (Kanungo & Dutta, 1966). In addition, the results lend strong support to the hypothesis that a fitting brand name derived from the association norms of the product is retained better than a nonfitting brand name. The fittingness variable seems to influence brand recall not only when brand names are low meaningful ($t = 7.07$, $p < .01$) but also when they are high meaningful ($t = 2.33$, $p < .05$). The meaningfulness variable, however, influences brand recall only when brand names are nonfitting ($t = 6.40$, $p < .01$) but not when they are fitting ($t = 1.67$, $p > .05$). Such interaction effects may interest the advertisers who want to choose between two HM or LM brand names. However, further replications are needed before the advertiser can place his confidence in such interaction effects.

In practice, many advertisers choose brand names that they consider intuitively as fitting to their products. Very often, such intuitive decisions are based upon one or a few individuals' remote associative responses to the products. As suggested in this study, a better method of choosing fitting brand names would be to use the conventional laboratory word-association techniques (Cofer, 1958) with a sample of potential consumers. Generally, the association techniques have been used for the purpose of studying language and verbal learning processes (Deese, 1959; Noble, 1952) and for clinical diagnoses (Jung, 1910; Kent & Rosanoff, 1910). The present study extends its usefulness to the selection of brand names by advertisers.

The pattern of recall for brand names of the three product-utility categories does not conform to the pattern of results reported by Kanungo and Dutta (1966). The authors reported that their male *S*s recalled fewer brand names of female-use products and their female *S*s recalled fewer brand names of male-use products than what was expected of them. These findings supported their hypothesis that brand awareness of high-utility products is superior to that of low-utility products. In the present study, however, both male and female *S*s recall male-use and female-use products equally well. These differences in the results of the two studies are presumably due to the difference in the culture patterns of *S*s. The *S*s used by Kanungo and Dutta (1966) were drawn from three Indian educational institutions where males and females live in a more or less segregated fashion. The cultural standards of India impose greater restrictions on free mixing among young unmarried male and female students compared to the North American cultural standards. Thus, for example, a female student in India would have less knowledge and concern about male-use products such as cigarettes or shaving blades. Likewise a male student in India would be less exposed to and interested in female-use products such as perfume or nail polish. On the other hand, the male and female *S*s used in the present study share each others interests to a greater extent due to greater exposure to each others interests in this culture. Thus it is not surprising that both male and female *S*s in the present study may be showing equal interest in both male-

use and female-use products. Hence they recall the brand names of male-use and female-use products equally well. These results therefore suggest the possible role of cultural factors in brand awareness, which need to be explored more fully in the future.

Finally, it may be recalled that a trend was noticed in both the male and female *Ss* to recall fewer brand names for the used-by-both products as compared to the brand names for either male- or female-use products. If it is assumed that the products in the used-by-both category, such as writing pads or adhesive tape, have less personal significance for both male and female *Ss* compared to the products in the other two categories, such as after shave lotion, nylons, or girdles, then the trend shown by the results are consistent with the hypothesis that brand awareness of high-utility products is better than that of low-utility products. The four products belonging to the used-by-both category may have been considered somewhat less important and therefore of lower utility than the products belonging to the other two categories. Unfortunately the products were not rated on a utility scale by the *Ss*. Such ratings might have substantiated the above explanation.

REFERENCES

COFER, C. N. Comparison of word associations obtained by methods of discrete single word and continuous association. *Psychological Reports*, 1958, **4**, 507–510.

DEESE, J. Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, 1959, **5**, 305–312.

JUNG, C. G. The association method. *American Journal of Psychology*, 1910, **21**, 219–269.

KANUNGO, R. N., & DUTTA, S. Brand awareness as a function of its meaningfulness, sequential position, and product utility. *Journal of Applied Psychology*, 1966, **50**, 220–224.

KANUNGO, R. N., & PANDA, K. C. Measurements of attributes of verbal materials: A further analysis of meaningfulness of CVC. *Journal of Psychological Research*, 1964, **8**, 131–135. (a)

KANUNGO, R. N., & PANDA, K. C. Some attributes of verbal materials: Evaluation of meaningfulness of some CVC triagrams. *Journal of Education and Psychology*, 1964, **22**, 26–30. (b)

KENT, G. H., & ROSANOFF, A. J. A study of association in insanity. *American Journal of Insanity*, 1910, **67**, 37–96, 317–390.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.

NOBLE, C. E. An analysis of meaning. *Psychological Review*, 1952, **59**, 421–430.

NOBLE, C. E. Measurements of association value (*a*), rated associations (*a'*), and scaled meaningfulness (*m'*) for the 2100 CVC combinations of the English alphabet. *Psychological Reports*, 1961, **8**, 487–521.

(Received August 7, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 4, 296–298

AGE, CIGARETTE SMOKING, AND TESTS OF PHYSICAL FITNESS

KENNETH H. DAVID [1]

University of Hawaii

It was hypothesized that smoking would decrease Ss' performance on physical fitness tests calling for strenuous activity, but not for tests calling for mild activity. 88 Ss were given 5 physical fitness tests as part of their work requirement and then asked to report the number of packages of cigarettes smoked per day. A multiple linear regression analysis of the results, using age as a covariate and amount of smoking as the independent variable, supported the hypothesis. The smoking conditions resulted in significant differences for the most strenuous test (1-mi. Run) and the 2 moderately strenuous tests (Crawling and Dodge & Jump), with performance scores decreasing as reported smoking scores increased. The least physically demanding tests showed no significant differences between smoking conditions. Age was an important factor in accounting for Ss' performance on all of the tests.

The present study was undertaken in order to investigate the relationship between varying amounts of cigarette smoking and several tests of physical fitness. If cigarette smoking affects an individual's performance on tests of physical fitness, the effect is probably mediated by the decreased efficiency of the circulatory and respiratory systems. It would appear probable that smoking reduces the oxygen capacity of the lungs; therefore, during physical activity the muscles may be deprived of sufficient oxygen, leading to feelings of fatigue and lowered performance. More strenuous activity would be expected to place a greater demand upon the circulo-respiratory system for oxygen than would mild activity. Thus, if smoking decreases the efficiency of a person's circulo-respiratory system, one would expect that smoking would decrease one's performance on tasks in which the circulo-respiratory system is of major importance, that is, tests in which vigorous physical activity is required. Conversely, one would expect that smoking would not affect one's performance on tests in which the circulo-respiratory system plays a minor role, that is, tests in which there is minimal physical activity.

In a study of the relationship between smoking and physical fitness, it was hypothesized that smoking would decrease *S*s' performance on tests calling for strenuous exercise, and that greater smoking would result in a greater decrease in performance. For fitness tests calling for mild physical activity, smoking was expected to have a negligible effect upon *S*s' performance. Thus, for the One-Mile Run test, which was the most strenuous test, smoking was expected to result in a greater decrease in performance than for any of the other tests. The Dodge & Jump test and the Crawling test were considered to be moderately strenuous, thus smoking was expected to result in decreased performance, but to a lesser degree than for the One-Mile Run test. The Climbing test was considered to be a mildly active test, thus smoking was expected to result in a slight or no decrease in performance. The Throwing test called for a minimal amount of activity, thus smoking was not expected to decrease performance.

METHOD

Subjects. The *S*s were 88 military personnel ranging in age from 19 to 39 yr. The physical fitness tests were administered to *S*s as part of their normal work requirement. During the 6 wk. preceding the testing, *S*s performed moderate calisthenics for 2 hr. per week. The number of calisthenics was the same for all *S*s, except that the older *S*s (30–39 yr.) performed 10% fewer exercises than the younger *S*s (19–29 yr.).

Tests. The five physical fitness tests were: (A) Crawling test—*S* was required to crawl for a distance of 120 ft., sliding along the ground at all times. The *S*'s score was the amount of time required to crawl the distance. (B) Climbing test—*S* was required to swing from rung to rung on a horizontal ladder without touching the ground. The *S*'s score was the number of rungs traversed during a 1-min. period. (C) Dodge & Jump test—*S* had to run across a

[1] Now at the University of Maryland, Far East Division.

TABLE 1

SUMMARY OF REGRESSION ANALYSIS OF SMOKING FOR FITNESS TESTS

Fitness test	M	SD	r^2		df	F
			Smoking and age	Age		
Crawling	31.59 sec.	6.28	.660	.596	5/81	3.07*
Climbing	45.18 rung	13.40	.348	.309	5/81	.97
Dodge & Jump	25.83 sec.	1.87	.564	.388	5/81	6.56**
Throwing	23.93 pt.	8.07	.526	.473	5/81	1.84
One-Mile Run	519.41 sec.	86.01	.644	.357	5/81	13.09**
Age	25.83 yr.	5.51				

* $p < .02$.
** $p < .001$.

49-yd. obstacle course and return to the starting position. The *S* had to avoid the obstacles and jump over a ditch during the running. The *S*'s score was the amount of time taken to complete the course. (D) Throwing test—*S* had to throw five grenades from a kneeling position, in which he attempted to hit the center of a target area 90 ft. away. The *S*'s score was based upon the combined scores of the five throws. Performance on this test is primarily a function of psychomotor skills, rather than demanding physical endurance. (E) One-Mile Run test—*S* was required to run around a 1-mi. grass track. The *S*'s score was the amount of time taken to run the distance. This task is a physically demanding test, in which physical endurance is of primary importance.

Procedure. The *S*s were separated into four groups of 22 *S*s each, assigned according to the alphabetical order of their last names. All four groups were given the Crawling test first and the One-Mile test last, with the other three tests given in a counterbalanced order. The order of the tests for the four groups was: ABCDE, ACBDE, ACDBE, and ADCBE. The *S*s were instructed to complete the four timed tests as quickly as possible and to strive for accuracy in the Throwing test.

After completing the tests, *S*s were required to write down the number of packages of cigarettes smoked per day, according to one of six choices: (*a*) none, (*b*) less than ½ per day, (*c*) ½ to less than 1, (*d*) 1 to less than 1½, (*e*) 1½ to less than 2, and (*f*) 2 or more. The number of *S*s falling within each category were: $a = 19$, $b = 14$, $c = 13$, $d = 14$, $e = 14$, and $f = 14$.

RESULTS AND DISCUSSION

The age of the *S*s appeared to be correlated with their performance scores; therefore, Pearson coefficients between age and performance were computed for the five physical fitness tests. The coefficients between age and the Crawling, Climbing, Dodge & Jump, Throwing, and the One-Mile Run tests were .772, −.556, .623, −.688, and .598 ($p < .01$, $df = 86$), respectively.

The results were analyzed by means of a multiple-regression analysis[2] (Bottenberg & Ward, 1963). A test for an interaction effect between age and the six smoking conditions was made for each of the physical fitness tests, with no significant differences occurring ($p > .05$). Since age was significantly correlated with the fitness tests, and age and smoking showed no significant interactions, age was used as a covariate with amount of smoking as the independent variable. The analysis showed that there were significant differences between the smoking conditions for three of the five physical fitness tests (see Table 1). The smoking variable showed significant differences for the Crawling test ($F = 3.07$, $df = 5/81$, $p < .02$), the Dodge & Jump test ($F = 6.56$, $df = 5/81$, $p < .001$), and the One-Mile Run test ($F = 13.09$, $df = 5/81$, $p < .001$), with no significant differences found for the Climbing test and Throwing test ($p > .05$).

In order to determine which of the smoking conditions were significantly different from each other, an analysis was made between each of the smoking conditions for the Crawling, Dodge & Jump, and One-Mile Run tests; thus, 45 comparisons were made using age as a covariate. A comparison of the six smoking conditions for the Crawling test showed that 11 of the 15 comparisons gave significant dif-

[2] Appreciation is expressed to Frank Ahern for the analysis and to the Computing Center of the University of Hawaii.

ferences ($p < .01$) in the predicted direction, while the other four comparisons showed no significant differences ($p > .05$). The Dodge & Jump test showed 13 significant differences between the smoking conditions, with the two nonsignificant differences occurring between conditions $\frac{1}{2}$–1 and 1–$1\frac{1}{2}$ packages ($p > .05$) and 1–$1\frac{1}{2}$ and $1\frac{1}{2}$–2 packages ($p > .05$). The significant differences were in the predicted direction, with greater smoking resulting in greater decrease in performance. The One-Mile Run test showed significant differences between all 15 comparisons of the smoking conditions ($p < .01$ for 14 of the comparisons, and $p < .02$ for the other). Thus, for the Crawling, Dodge & Jump, and One-Mile Run tests, an increase in self-reported smoking was associated with a decrease in performance.

The analysis of the results supported the hypotheses. The smoking conditions had a highly significant effect upon the most strenuous of the physical fitness tests, the One-Mile Run test, with smoking accounting for approximately 29% of the variance. Smoking also showed significant effects for the two moderately strenuous tests, with the Dodge & Jump and Crawling tests accounting for approximately 18% and 6% of the variance, respectively. Smoking showed no significant effects for the two least physically demanding tests, the Climbing and Throwing tests. Therefore, the results support the position that smoking results in decreased performance of physical fitness tests which require strenuous activity, and with a decrease in the required endurance of the test there is a corresponding decrease in the effect of smoking.

The age factor was of greater importance than smoking in accounting for the differences between *S*s' scores on the physical fitness tests (see Table 1). For the Crawling, Climbing, Dodge & Jump, Throwing, and One-Mile Run tests, the age factor accounted for approximately 60%, 31%, 39%, 47%, and 36% of the variance ($p < .001$), respectively.

REFERENCE

BOTTENBERG, R. A., & WARD, J. H., JR. Applied multiple linear regression. Technical Report PRL-TDR-63-6, Project 7719, Task 771901. Washington, D. C.: Clearinghouse for Federal Scientific and Technical Information, United States Department of Commerce, 1963.

(Received August 9, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 4, 299-303

ABILITY, VALUES, AND CONCEPTS OF EQUITABLE SALARY INCREASES IN *EXERCISE COMPENSATION*

BERNARD M. BASS

Management Research Center, University of Rochester

113 American graduate business students recommended annual salary increases for 10 fictitious engineers, each with a different reason for being given a raise. The mean recommendation was 9.3%. Recommenders were more generous if the students were of lower intelligence and achievement, and if they held strong social and service rather than theoretical and economic values. They also were more willing to spend company money for other social and personnel purposes.

Several recent reviews point to increasing interest in psychological aspects of pay (Adams & Jacobsen, 1964; Haire, Ghiselli, & Porter, 1963; Weick, 1966). A modest amount of available research has probed the effects of pay differentials on attitudes and performance of those receiving the money, but little has been done to ascertain what motivates those who decide on what salaries will be paid when they respond to requests for salary increases and how their own ability and values affect their monetary decisions. These latter issues are the concern of this report.

METHOD

Subjects

Forty-four percent of the 113 male graduate business students who participated in this study held undergraduate engineering degrees, the other 56% had earned undergraduate degrees in the liberal arts or in business administration. They averaged close to "B" in undergraduate grades and earned a mean of 566 on the Admissions Test for Graduate Schools of Business (ATGSB).

Procedure

During the fourth week of a course in behavioral science, each participant was involved in a series of small group exercises (Bass, 1965, 1967). One of these, Exercise Compensation (First ed.), required that each participant, by himself, make a set of 10 salary recommendations prior to a group discussion about the matter. He was given the instructions which in essence were as follows:

> You will have to recommend increases for 10 engineers in your company, all of whom are unmarried, 22-year-old college graduates who have just completed their first year with the company and are now to be considered for an annual raise. Keep in mind that you may be setting precedents, that you need to keep down salary costs, that you need to consider equity, that you want to retain all the engineers if possible. Your company has no job evaluation schedule but tries to match community and industry norms and to be competitive.
>
> The 10 engineers are as follows:
>
> *Al*, an average performer (50th percentile);
> *Bill* (90th percentile in merit);
> *Charlie* (10th percentile in merit).
>
> All the remainder are like Al in merit, but in addition:
>
> *Dan* is working on an insecure government contract. Because of his specialty, transfer is difficult.
> *Ed* is on a "dead-end" job lacking advancement potential.
> *Frank* has a dirty, unpleasant, hazardous, uncomfortable job.
> *Garry* has a low-prestige job in a low-prestige subsidiary.
> *Henry* has to work closely with a crew of uncongenial, unfriendly, competitive co-workers.
> *Irwin* has a boring, dull, and monotonous job.
> *Jim* has a definite job offer from a competitor substantially better than his current salary.

Other Measures

Of the 113 participants, during other weeks of the course, 112 completed the Allport-Vernon Study of Values and ranked their own life goals on the first edition of Exercise Life Goals (Bass, 1965). The goals were defined as follows:

(*a*) Leadership: To become a community leader, to become influential in public affairs, to become influential among co-workers.

(*b*) Comfort: To have the time and means to enjoy life, to become happy and content, to have financial security, to be well-liked.

(*c*) Creativity: To produce original and unique plans and ideas, to devise new approaches, to be expressive.

(*d*) Expertness: To become an authority on a special subject in any field, to persevere to reach a hoped-for expert level of skill and accomplishment.

(*e*) Technology: To make contributions to technical or scientific knowledge, to invent apparatus or equipment.

TABLE 1

RELATION BETWEEN TESTED INTELLIGENCE ACHIEVEMENT AND COMPENSATION DECISIONS

Recommendations for	ATGSB			Undergraduate grades[b]	Mean % increase in recommendations[c]	*SD* of recommendations[c]
	L[a]	Q[a]	T[a]			
Al (average)	−.08	−.18*	−.17*	−.19*	8.02	2.62
Bill (best)	−.08	−.14	−.15	−.25**	11.96	3.78
Charlie (poor)	−.17*	−.17*	−.23*	.00	5.55	2.38
Dan (no security)	−.09	−.11	−.15	−.03	9.83	3.24
Ed (no opportunity)	−.12	−.21*	−.22*	−.09	9.05	3.00
Frank (poor conditions)	−.01	−.24**	−.16*	−.04	10.18	3.08
Garry (no prestige)	−.10	−.21*	−.20*	−.14	9.49	3.16
Henry (unfriendly co-workers)	−.05	−.21*	−.18*	−.13	9.72	3.10
Irwin (uninteresting work)	.00	−.19*	−.12	−.14	9.83	3.26
Jim (competitive offer)	−.02	−.16*	−.11	−.08	9.34	3.32

Note.—ATGSB = Admissions Test for Graduate Schools of Business. L = Linguistic; Q = Quantitative; T = Total.
[a] $N = 101$.
[b] $N = 99$.
[c] $N = 113$.
* $p < .05$.
** $p < .01$.

(*f*) Prestige: To become well-known, to obtain awards and recognition.

(*g*) Service: To contribute to the satisfaction of others, to be helpful to others who are in difficulty.

(*h*) Wealth: To earn a great deal of money, to build a large financial estate.

(*i*) Independence: To be free to do what one likes, to be one's own boss, to control others rather than be controlled by them.

Exercise Objectives (Bass, 1965) also was completed by the 113 participants. In this activity, after reading a company's last annual profit and loss statement, each participant made five budgeting decisions. In one of these he had the option to spend or not to spend $225,000 to correct a safety hazard likely to produce a serious injury. In another, he could spend or avoid spending $150,000 to modify a plant that was polluting a stream. In still others, he had to spend money on managers, products, and to settle a strike, but in these cases his problem was on which way to spend the money to handle each case. These latter decisions were not expected to relate to salary recommendations and in fact they did not do so.

RESULTS

Mean Recommendations

The mean increase awarded the 10 engineers by the 113 graduate business students (Table 1, column 5) was 9.3%. As is common with most other groups of managers and students tested here and abroad, the largest source of variance in recommendations was merit. Bill, of high merit, was assigned 1.5 times the increase awarded Al, the average man, while Charlie, the man of low merit, was given approximately .7 as much as average Al. The hypothesis has been advanced (Bass, 1967) that half of the increase given Al, the average man, will be denied to Charlie. This half will be added to Bill's increase so that Bill gets 1.5 times as much as Al, and Charlie is left with only .5 of the increase given Al. But the hypothesis did not hold completely in Charlie's case. Here, the ratio of mean recommendations that occurred for Bill:Al:Charlie was 3:2:1.5, rather than 3:2:1.

As seen in Table 1, column 5, next to meritorious performance the participants were most willing to pay for poor working conditions (Frank) and were relatively least responsive to Ed who lacked opportunities, although in comparison to what they awarded average Al, they were willing to give a premium for everyone from Dan to Jim who had special problems.

Correlational Analyses

The main purpose of this report is to see what personal factors influenced the recommendations. For this, correlational analyses will be presented in four sections. First, the extent that salary recommendations are associated with the intelligence and achievement

of the recommender will be examined. Next, will be seen how the self-reported values of the recommender may enter the picture. Third, a look will be taken at how the recommender's own stated life goals correlated with his recommendations for raises in salary. Finally, the author will note how generosity in recommending raises also may relate to willingness to spend company money on other personnel and organizational needs.

Ability and Compensation Decisions

Table 1 shows the product-moment correlations between the salary increases each participant was willing to award each of the 10 engineers and each participant's tested intelligence and undergraduate record (columns 1–4). It can be seen that differences among the participants' decisions were unrelated to verbal aptitude as measured by the ATGSB, except possibly the salary decisions about Charlie, who was lacking in merit. However, a modest association was found between quantitative aptitude and the tendency to recommend smaller raises in salary. Participants with greater tested quantitative aptitude awarded smaller salary increases, particularly ($p < .01$) when the increases were being justified by poor working conditions ($r = -.24$). When verbal and quantitative scores were combined and total intelligence test scores were correlated with compensation decisions, the effect was most pronounced in dealing with the salary increase for Charlie, the engineer low in merit. Those with greater tested intelligence were significantly likely to recommend a smaller raise for Charlie ($r = -.23$).

Perplexing was the significant correlation ($p < .01$) of $-.25$ between undergraduate quality-point averages and the raise awarded Bill, suggesting that achieving students recommended smaller salary increases for the most meritorious engineer. They awarded Al, the average engineer, a smaller raise also. There was little other evidence of the influence of undergraduate achievement on the other compensation decisions.

To sum up, some tendency was seen for brighter and more academically successful participants to recommend smaller raises (or for the duller and less successful to recommend bigger salary increases). Quantitative aptitude was more of a factor than verbal aptitude.

Values, Life Goals, and Compensation Decision

The correlations among the Allport-Vernon Study of Values scale scores and the salary recommendations are shown in Table 2. It can be seen that holding certain values predisposed recommenders to be generous, while strong values in other directions did the reverse. Those with strong social values tended to recommend bigger raises in general to all engineers, but the effects were less marked

TABLE 2

RELATION BETWEEN ALLPORT-VERNON STUDY OF VALUES AND COMPENSATION DECISIONS

Recommendations for	Allport-Vernon scores					
	Theoretical	Economic	Aesthetic	Social	Political	Religious
Al (average)	−.23**	−.38*	.10	.34**	−.03	.16*
Bill (best)	−.13	−.18*	−.09	.20*	.07	.11
Charlie (poor)	−.20*	−.38*	.08	.33**	−.06	.17*
Dan (no security)	−.14	−.20*	−.10	.24**	.11	.10
Ed (no opportunity)	−.23**	−.27**	−.01	.33**	.04	.14
Frank (poor conditions)	−.16*	−.26**	−.05	.19*	.09	.19*
Garry (no prestige)	−.29**	−.27**	.03	.32**	.04	.16*
Henry (unfriendly co-workers)	−.24**	−.29**	.04	.29**	.08	.12
Irwin (uninteresting work)	−.30**	−.30**	.05	.29**	.09	.17*
Jim (competitive offer)	−.20*	−.24**	.02	.17*	.07	.17*

Note.—$N = 112$.
* $p < .05$.
** $p < .01$.

TABLE 3

RELATION BETWEEN SELF-RANKED LIFE GOALS AND COMPENSATION DECISIONS

Recommendations for	Self-ranked life goals								
	Leadership	Comfort	Creativity	Expertness	Technology	Prestige	Service	Wealth	Independence
Al (average)	.10	−.07	−.11	−.10	−.09	.02	.26**	−.03	.01
Bill (best)	.15*	−.14	−.00	−.13	−.03	.09	.22**	−.06	−.10
Charlie (poor)	.09	−.11	−.09	−.00	−.12	−.00	.18*	−.03	.02
Dan (no security)	.10	−.08	−.14	−.11	−.19*	.06	.27**	−.00	.04
Ed (no opportunity)	.12	−.09	−.15*	−.07	−.19*	−.01	.29**	.02	.02
Frank (poor conditions)	.12	−.07	−.11	−.07	−.13	−.00	.31**	−.04	−.04
Garry (no prestige)	.11	−.12	−.10	−.13	−.16*	.03	.27**	.04	−.01
Henry (unfriendly co-workers)	.09	−.06	−.06	−.12	−.18*	.03	.27**	.02	.06
Irwin (uninteresting work)	.07	−.05	−.06	−.14	−.20*	.02	.27**	−.00	.04
Jim (competitive offer)	.15*	−.03	−.15*	−.16*	−.21*	.03	.27**	.02	.02

Note.—Signs reversed. $N = 113$.
* $p < .05$.
** $p < .01$.

when dealing with Bill (most meritorious), Frank (poor working conditions), and Jim (competitive offer). Evidently, lack of security, opportunity, prestige, congenial co-workers, and interesting work weighed heavier as an excuse to grant bigger raises for recommenders scoring high in social values.

The correlations also were uniformly positive between religious value scores and the size of salary recommendations, but they were generally lower. Strong social values played a greater role than strong religious values. At the other extreme, recommenders holding strong economic and theoretical values according to their self-reports were likely to offer smaller raises to the engineers. Those holding strong economic values were particularly unlikely to be generous to Al (average merit) and Charlie (low merit). Those holding strong theoretical values were particularly unlikely to be generous to George (no prestige) and Irwin (unfriendly co-workers).

TABLE 4

RELATION BETWEEN WILLINGNESS TO SPEND MONEY ON SAFETY AND STREAM CLEANING AND COMPENSATION DECISIONS

Engineer	Spend on safety	Spend on stream clean-up
Al (average)	.07	.09
Bill (best)	.10	.14
Charlie (poor)	.09	.11
Dan (no security)	.23*	.11
Ed (no opportunity)	.20*	.13
Frank (poor conditions)	.18*	.11
Garry (no prestige)	.21*	.11
Henry (unfriendly co-workers)	.21*	.13
Irwin (uninteresting work)	.19*	.16*
Jim (competitive offer)	.22**	.13

Note.—$N = 113$.
* $p < .05$.
** $p < .01$.

Table 3 shows how self-ranked life goals were related to compensation recommendations. Here one goal stands out. Recommenders with relatively greater concern for service as an important life goal tended to be more generous—almost uniformly so.

To a lesser extent, those concerned about technical success were likely to be less generous than the average recommender.

Different reasons for being generous may or may not be involved. For those high in economic value, it may be that the same amount of money is worth more and is considered to be a larger raise than would be seen by those low in economic value. For those high in concern for theory, understanding, and technical success, less need may be seen to grant large increases in compensation

for effort. For those with strong concerns for social service and religion, pure generosity may be operative in their salary recommendations.

Willingness to Spend Money

How general is generosity? Two of the five budgeting decisions on Exercise Objectives involved whether or not to spend money (not how to spend it) primarily for human or social rather than immediate economic returns. As can be seen in Table 4, there was a slight positive association between the size of salary increases recommended and a willingness to spend money on fixing a plant to end pollution of a stream. There seemed to be an even greater tendency to provide extra money for raises for all kinds of special reasons other than merit (r's ranged from .18 to .23) and a willingness to spend money on safety equipment to avoid the possibilities of a serious accident.

CONCLUSIONS

Despite the observation that "Pay, in one form or another, is certainly one of the mainsprings of motivation in our society" (Haire et al., 1963), little data are available on what bases decisions about pay are made. Many people would like to believe that highly objective analyses produce highly objective job evaluations and merit rating and manpower plans which yield highly objective standards for establishing pay and pay raises. In fact, the author suggests that much subjectivity of judgment is involved at every phase in these schemes and that it would be useful to develop some understanding of what may affect these judgments. In the present study, one sees that while merit overshadowed other factors in consideration by the recommenders of salary increases, at the same time, as might have been anticipated, the abilities and values of the recommenders played some role in what values they were willing to recommend.

Higher salary increases were recommended by those lower in intelligence and achievement, by those who held strong social and religious rather than theoretical or economic values, who were more interested in service rather than technical achievement and who also were more willing to spend their company's money for other personnel and social purposes.

REFERENCES

ADAMS, J. S., & JACOBSEN, P. R. Effects of wage inequities on work quality. *Journal of Abnormal and Social Psychology*, 1964, 69, 19–25.

BASS, B. M. *A program of exercises in management and organizational psychology*. Pittsburgh: Management Development Associates, 1965.

BASS, B. M. Combining management training and research. *Training Directors Journal*, 1967, 21(4), 1–7.

HAIRE, M., GHISELLI, E. E., & PORTER, L. W. Psychological research on pay: An overview. *Industrial Relations*, 1963, 3, 1–8.

WEICK, K. E. The concept of equity in the perception of pay. *Administrative Science Quarterly*, 1966, 11, 414–439.

(Received August 25, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 4, 304–308

RELATIONSHIP BETWEEN PERCEPTUAL STYLE AND SIMULATOR SICKNESS

GERALD V. BARRETT AND CARL L. THORNTON

Goodyear Aerospace Corporation, Akron, Ohio

Simulator sickness was hypothesized to be caused by the conflict between the visual presentation of apparent motion and the lack of any corresponding body sensation of motion. The hypothesis was tested by correlating individual differences in scores on the Rod and Frame Test (RFT; which measures accuracy of adjustment of a rod to true vertical under conditions of visual-kinesthetic conflict) and degree of simulator sickness. The data for Series 3 of the RFT and the indexes of sickness were best represented by hyperbolic functions yielding correlations of .40–.52. Implications for simulation technology and for a general conflict of cue theory are discussed with emphasis on supporting evidence from several areas of investigation.

Since World War II, simulators have been developed which give the visual illusion of motion without any actual physical motion. Unfortunately, certain *S*s became ill while operating these devices. The illness phenomenon was first intensively studied by Miller and Goodson (1958, 1960), who labeled it motion sickness because the symptoms resembled those experienced by some people in moving vehicles. There was, however, no reason to prematurely so label this phenomenon motion sickness since motion is not involved in fixed-base simulators. As Tyler and Bard (1949) have pointed out, the primary cause of motion sickness is probably motion, and the failure to appreciate this fact can lead to confusion in conceptualization. In this paper the term "simulator sickness," rather than motion sickness, will be used to denote the symptoms which occur in fixed-base simulators incorporating a moving visual scene.

A number of hypotheses have been informally advanced to explain simulator sickness: distortion of vertical objects, rapid change in brightness, too much detail, poor resolution, excessive lag between simulator controls and corresponding shift in visual display, high-frequency vibrations which disrupt accommodation, distance between the visual display and the observer such that accommodation is different from that usually experienced, and conflict between the apparent motion seen on the visual display and lack of any corresponding motion of the simulator. This last hypothesis was investigated in the present study.

Several other conflict situations have been found to produce sickness. Wood (1895) described an amusement park device with a large immobile swing inside a movable room. When the room moved many individuals experienced considerable discomfort. Crampton and Young (1953) induced nausea in *S*s seated in a fixed chair in the center of a rotating room. Discomfort lasted for some time after the experience, for one *S* up to 2 days. This wide range of sickness time (from 0 to 48 hr.) is noteworthy, since all *S*s were subjected to the same experience.

While conflict between visual and body cues may be the dominant cause of simulator sickness, other factors such as type of simulator, fidelity of simulation, *S*'s experience, and *S*'s involvement are also important. Fixed-base simulators may have two types of visual displays: outside-in (O-I), such as a child's remote control car, or inside-out (I-O), in which the operator views the scene as he would from inside a real vehicle. While sickness has been quite common with an I-O display, none has been reported for O-I simulators. Low involvement resulting from being unable to put oneself psychologically into the vehicle is probably the reason that O-I displays do not cause sickness. A comparison of I-O and O-I displays (Matheny, Dougherty, & Willis, 1963) revealed that performance improved with an I-O display (but not O-I) when motion cues were added. Since the op-

erator "expects" motion in an I-O situation, cue conflict which leads to sickness and/or decrement in performance can occur when such expected motion is not present.

Concerning *S*'s experience, Miller and Goodson (1958) found that 60% of the experienced helicopter pilots but only 12% of the student pilots tested reported simulator sickness. There was probably no cue conflict for the student pilots since they had not learned the particular body inputs which occur during helicopter operation. Fitts (1951) suggested that visual control was important when an individual first learns a perceptual-motor task, and proprioceptive feedback or "feel" becomes more important as experience increases. Fleishman and Rich (1963), for example, using predictors of a perceptual-motor task, found that a visual test correlated significantly with the task in early but not later trials. Conversely, a kinesthetic test correlated on the later but not the early trials. This indicates that initially visual cues guide performance, but with experience individuals sensitive to proprioceptive cues can become most proficient.

The importance of body cues was also demonstrated by a factor analysis of 12 driving tests (Herbert, 1963). One of the main factors was labeled proprioception because of the high loading of a nonvisual driving task. Experience with the necessary proprioceptive feedback, then, appears to be necessary before cue conflict can result.

Fidelity of simulation would be the degree to which the simulated conditions approach conditions of the real world. If the simulator were of low fidelity no conflict or sickness would result since *S*s cannot become involved in the very unrealistic task.

The *S*s will also not become involved in the task if they have a "play set." Their behavior will not be pertinent to good driving performance with, for example, an attempt to crash the simulated vehicle. Low involvement yields no conflicting cues and therefore no simulator sickness. This conclusion is consistent with the finding that people do not become ill in amusement park simulators even though all the other necessary conditions may be present.

Recently Barrett and Nelson (1965d, 1966b) evaluated an automobile simulator which had all the aforementioned parameters: high fidelity, I-O display, and experienced and involved *S*s. They found symptoms which were quite similar to those reported in previous simulator research, including cold sweating, upset stomach, vertigo, dizziness, nausea, feeling of faintness and disorientation. About half of the *S*s became too ill to continue after only 5–10 min. Two *S*s became so ill as to regurgitate.

A research program was initiated to test the cue conflict and conflict sensitivity hypothesis, as measured with a Rod and Frame Test (RFT) apparatus following the field-dependence conceptualization of Witkin, Lewis, Hertzman, Machover, Meissner, and Wapner (1954). Field-independent *S*s were deemed more sensitive to body cues than field-dependent *S*s. On this basis it was predicted that the field-independent *S*s would experience more discomfort in the simulated situation than field-dependent *S*s.

METHOD

Automobile Simulator

An unprogrammed automobile simulator was the basic research tool. A terrain model, an 87:1 (HO gauge) scale representation of several flat roads, supplied the visual scene. Mounted above the terrain model was a television camera with motions in direct response to the movements of the brake, accelerator, and steering wheel of the automobile. Thus, *S* had complete control over the part of the terrain model that the camera traversed.

A projected image visual display was used with half of the *S*s and a virtual image display with the remainder. Both visual displays gave the driver approximately a 50° horizontal angle view of the terrain model and a center resolution of approximately 500 lines. There was no significant difference between the displays in the percentage who left the simulator because of discomfort. Therefore, the two displays were considered to be functionally equivalent. With each visual display the driver sat in the automobile and performed the usual control movements associated with driving a car. A more detailed description of the simulator and associated visual displays has been reported by Barrett and Nelson (1965c, 1966a).

Perceptual style was measured with a standard RFT (Witkin et al., 1954). The apparatus consisted of a luminous frame (40 in. sq.) pivoted at its center so that it could be tilted left or right. Pivoted at the same center, but moving independently of

TABLE 1

RELATIONSHIP BETWEEN PERCEPTUAL STYLE AND DISCOMFORT

Measure of discomfort	Rod and frame measures			
	$1/S_1$	$1/S_2$	$1/S_3$	1/OI
Discomfort	.14	.13	.40*	.23
Illness after	.22	.31	.52**	.44*
No. trials	−.22	−.11	−.45**	−.25
Stayed left	−.18	−.14	−.47**	−.26

Note.—S_1 = Series 1; S_2 = Series 2; S_3 = Series 3; OI = Orientation Index.
* $p < .01$.
** $p < .001$.

the frame, was a luminous rod (39 in.). The *S* was seated 8 ft. from the rod and frame in a chair which could be placed in three positions: erect, tilted 28° to the left, or tilted 28° to the right. Testing was done in a completely darkened room and *S* wore goggles with dark lenses so that he could see only the rod and frame.

Subjects

A random sample of 50 male *S*s between the ages of 30 and 45 were selected from approximately 1,200 employees in an engineering division of an aerospace corporation. Approximately 6 mo. after the completion of the simulator evaluation, 46 *S*s (23 from each display) were able to be recontacted and tested with the RFT.

Procedure for Automobile Simulator Investigation

Each *S* drove three orientation trials around the terrain model followed by a pretrial run for a study of driving at requested speed. After making the tenth and final speed judgment, *S* was exposed to an emergency situation where he had to stop for a suddenly emerging pedestrian dummy. The procedures have been described in greater detail by Barrett and Nelson (1965a, 1966b). During the evaluation of the simulator, *S* was observed by the experimenter. If *S* complained of discomfort he was told that he was free to leave at any time.

RFT Procedure

Approximately 6 mo. after the data emergency behavior study, the perceptual style of *S*s was measured following the standard procedure (Witkin et al., 1954). Series 1 (S_1) of the RFT consisted of eight trials in which *S* and the frame were tilted 28° in the same direction; Series 2 (S_2) were eight trials in which *S* and frame were tilted 28° in opposite directions; Series 3 (S_3) consisted of eight trials with the frame tilted 28° to the right or left while *S* remained upright. The *S*'s task was to position the rod to what he considered to be true vertical by asking the experimenter to move the top of the rod right or left. The *S*'s score was the number of degrees in error in each series. In addition, an Orientation Index (OI) was computed from the standard scores of the three series.

RESULTS

Four measures of discomfort were compared to RFT scores. The first measure was *S*'s rating of discomfort, using a 0–10 graphic rating scale. The scale was part of a 10-item questionnaire concerning the simulator which was administered 6 mo. after the simulator study. A second questionnaire measure was *S*'s estimate of the length of time after leaving the simulator that the discomfort persisted—termed "illness after." Responses to this question ranged from zero illness to 48 hr. Third, the number of trials *S* was able to remain in the simulator was used as an index of discomfort. The range was from 1/4 trial to completion (14 trials). Fourth, *S*s were divided into two categories: those who completed all trials and those who did not. Twenty-three of 46 *S*s were able to complete the simulator study.

Linearity was approximated by reciprocal transformations of the perceptual style measures. Correlations are shown in Table 1 where it can be seen that the only consistently significant relationships were between S_3 and the four measures of sickness.

Table 2 shows an apparent threshold phenomena. The *S*s were classified according to an adult standardization sample for S_3 (Witkin et al., 1954) with those who were either 1 standard deviation above or below the mean labeled extreme field dependent or extreme

TABLE 2

RELATIONSHIP BETWEEN LEAVING SIMULATOR AND PERCEPTUAL STYLE CLASSIFICATION ON SERIES 3

Subject	Extreme field independent	Field independent	Field dependent	Extreme field dependent
Left simulator (N = 23)	12	6	2	3
Remained in simulator (N = 23)		21		2

field independent. All the extremely field-independent *Ss* left the simulator.

In order to determine if other aspects of the RFT test would significantly add to the relationship, multiple-regression equations were calculated using x, x^2, $1/x$, in x-data transformations of S_1, S_2, S_3, and OI. No measures were found to add significantly to the variance accounted for by S_3.

Mention should be made of the possible reason for S_3 being the only measure which was related to simulator discomfort. Witkin et al. (1954) statistically analyzed 10 perception subtests and found three general groupings. S_3 was not in the same grouping of perception tests as were S_1 and S_2, indicating that the RFT taps at least two perceptual factors and that only one of them was related to simulator sickness.

DISCUSSION

If the conflict of cue hypothesis is correct, an obvious remedy to the simulator sickness problem is to introduce a certain degree of physical motion into the simulated system. An interesting question is the degree of motion required to give the necessary body cues. Simple random vibration may be enough to eliminate the cue conflict, a possibility having considerable practical and economic import for the simulation art.

Besides the specific cue conflict when motion is lacking, the results may have implications for motion sickness research, in which case motion is present with inappropriate or missing visual cues. This is supported by the finding that those who experience motion sickness were also likely to experience sickness in a conflict situation (Crampton & Young, 1953). It has been reported (Clark, 1963) that some pilots become disoriented while flying in extreme haziness or cloud covering. They may be receiving adequate motion cues but not the corresponding visual inputs. In much the same manner, while flying under instrument conditions (again motion but no vision), some pilots become disoriented and mistrust their instruments.

Simulation of zero gravity in aircraft produced illness for 50–70% of the *Ss* tested by Loftus (1963). It is possible that cue conflict experienced in this unusual situation induced the illness. The finding that labyrinthine-defective *Ss* (fewer body cues) showed no signs of zero gravity sickness, while 64% of the normal *Ss* did, supports this hypothesis (Kellogg, Kennedy, & Graybiel, 1964).

Wendt (1951) concluded that the vestibular apparatus per se causes motion sickness since people with no vestibular sensitivity do not become sick. An alternate conclusion might be that, since labyrinthine-cue-related body sensitivity is low for deaf people, it is less likely that they will experience any conflict in cues.

It is possible that both cue conflict and vestibular stimulation are important, but under different conditions. Walsh (1962) oscillated *Ss* in the horizontal plane. At 1 cps the *Ss* correctly felt they were traveling in a given direction. When the oscillations were 1/3 cps or slower, the sensations of moving were in anticipation of the motion, there being a phase advance. Sensations of motion in one direction were frequently aroused when the person was still traveling rapidly in the opposite direction, thus conflict. It is possible that motion sickness may have different causes depending upon the rate of motion. At slow rates the conflict of cues may apply; with greater accelerative forces the associated symptoms might be due to excessive stimulation of the vestibular mechanism.

The fact that pilots and drivers rarely become sick and passengers often do (Tyler & Bard, 1949) can be explained by the conflict of cue hypotheses. Since the driver receives direct feedback from the vehicular controls, it is understandable that he does not experience conflict, while the passenger has no such referent and may become ill.

The question remains as to why certain field-dependent people became ill in the simulator situation. Other variables such as the physical condition of *Ss* may have had some influence. Another explanation can be found from early perceptual style research. Witkin et al. (1954) stated that some *Ss* experienced great difficulty in making a judgment on the Body Adjustment Test. They appeared to be influenced by the visual scene but also aware of body position. Being unable to consistently utilize body position, their responses were quite variable. They eventually become dis-

oriented, with some experiencing physical discomfort akin to simulator sickness. The field-dependent *Ss* who function in this manner may also experience simulator sickness. To probe this possibility the 10 most-field-dependent *Ss* were compared as to variability of responses. Of the five sick *Ss*, four were extremely variable (range = 20°) and four of the five nonsick *Ss* were very consistent in their responses (range = 5°). While the results were suggestive, a large subsample would be required for statistical confirmation.

Although cue sensitivity in terms of perceptual style appears to explain the extreme discomfort that some people experience, Witkin, Dyk, Faterson, Goodenough, and Karp (1962) recently discussed perceptual style in terms of being able to extract an item from an embedded context. The results of the study could be explained in these terms also. In a simple laboratory study of kinesthetic sensitivity where cues are isolated there may be no differences between field-dependent and field-independent *Ss*. However, in the complex simulator (an embedding context) the field-independent person may be more aware of the cues which are in conflict (i.e., can disembed them), and thus he becomes ill.

REFERENCES

BARRETT, G. V., & NELSON, D. D. Driving at requested speed. Volume 12, Goodyear Engineering Report-12400, Contract No. PH-108-64-168, December 1965. (a)

BARRETT, G. V., & NELSON, D. D. Emergency behavior. Volume 13, Goodyear Engineering Report-12400, Contract No. PH-108-64-168, December 1965. (b)

BARRETT, G. V., & NELSON, D. D. Introduction to driving simulator evaluation. Volume 1, Goodyear Engineering Report-12400, Contract No. PH-108-64-168, December 1965. (c)

BARRETT, G. V., & NELSON, D. D. Summary of human factors evaluation. Volume 15, Goodyear Engineering Report-12400, Contract No. PH-108-64-168, December 1965. (d)

BARRETT, G. V., & NELSON, D. D. Description of virtual image display. Volume 16, Goodyear Engineering Report-12400, Contract No. PH-108-64-168, March 1966. (a)

BARRETT, G. V., & NELSON, D. D. Summary of human factors evaluation with virtual image display. Volume 20, Goodyear Engineering Report-12400, Contract No. PH-108-64-168, March 1966. (b)

CLARK, B. Visual space perception as influenced by unusual vestibular stimulation. *Human Factors*, 1963, **5**, 265–274.

CRAMPTON, G. H., & YOUNG, F. A. The differential effect of a rotary visual field on susceptibles and nonsusceptibles to motion sickness. *Journal of Comparative and Physiological Psychology*, 1953, **46**, 451–453.

FITTS, P. M. Engineering psychology and equipment design. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1287–1340.

FLEISHMAN, E. A., & RICH, S. Role of kinesthetic and spatial-visual abilities in perceptual-motor learning. *Journal of Experimental Psychology*, 1963, **66**, 6–11.

HERBERT, M. J. Analysis of a complex skill: Vehicle driving. *Human Factors*, 1963, **5**, 363–372.

KELLOGG, R. S., KENNEDY, R. S., & GRAYBIEL, A. Motion sickness symptomatology of labyrinthine defective and normal subjects during zero gravity maneuvers. Aerospace Medical Research Laboratory, Tech. Documentation Rep.-64-47, June 1964.

LOFTUS, J. P. Motion sickness in the C-131B. In J. P. Loftus (Ed.), Symposium on motion sickness with special reference to weightlessness. Aerospace Medical Research Laboratory, Tech. Documentation Rep.-63-25, June 1963. Pp. 3–5.

MATHENY, W. G., DOUGHERTY, D. J., & WILLIS, J. M. Relative motion of elements in instrument displays. *Aerospace Medicine*, 1963, **34**, 1041–1046.

MILLER, J. W., & GOODSON, J. E. A note concerning motion sickness in the 2-FH-2 hover trainer. Naval Aviation Medical Center, Pensacola, Florida, Project NM 17-01-11 Subtask 3, Report No. 1, February 1958.

MILLER, J. W., & GOODSON, J. E. Motion sickness in a helicopter simulator. *Aerospace Medicine*, 1960, **31**, 204–211.

TYLER, D. B., & BARD, P. Motion sickness. *Physiological Review*, 1949, **29**, 311–369.

WALSH, E. G. The perception of rhythmically repeated linear motion in the horizontal plane. *British Journal of Psychology*, 1962, **53**, 439–445.

WENDT, G. R. Vestibular functions. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1191–1223.

WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., & KARP, S. A. *Psychological differentiation: Studies of development*. New York: Wiley, 1962.

WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.

WOOD, R. W. The "haunted swing" illusion. *Psychological Review*, 1895, **2**, 277–278.

(Received August 28, 1967)

Journal of Applied Psychology
1964, Vol. 52, No. 4, 309–312

EVALUATION OF A NEW HYBRID IN PROGRAMMED INSTRUCTION

D. H. SCHUSTER [1]

Iowa State University

A new format of programmed instruction similar to the branching style of Crowder was developed for the subject matter of electronic test equipment. All of the answers were explained on the page following a question. This hybrid was evaluated and compared with the conventional scrambled format in a field tryout with electronic technicians. There was no significant difference in achievement between the 2 formats, whether additional participation by the use of preperforated punch cards was considered or not. There was a tremendous opinion difference in favor of the linearized version afterward, regardless of whether the student had actually taken the linear format or not.

The aim of this research was to try out with electronic technicians a new format in programmed instruction and compare it with a conventional scrambled book or Crowderian format. A secondary question was to evaluate the effectiveness of additional participation in the program by means of overt responses.

One type and format of programmed instruction, due to Crowder (1960), is the following: A page of instructional material is ended by a question with several multiple-choice answers. Only one of these choices is correct. The choices are answered and discussed on separate pages distributed at random throughout a given chapter. If a student selects a wrong answer, the wrongness of that particular answer is discussed in view of the instructional material and the student is usually returned to the original instructional page to reread it and to try again. If a student selects the correct answer, he is told that this is the correct answer and an explanation is given. Then the student goes on to the next page of instructional material with its terminating question and so on through a chapter. A book constructed this way is said to be "scrambled" or "self-tutoring."

From a preliminary study with a scrambled book on electronic test equipment, approximately half of the students stated that they would like to see the scrambled book unscrambled or linearized. The writer accordingly modified the scrambled book program such that all of the answers, correct and incorrect, were listed and explained on the page immediately following the instructional page with its ending question. Thus the student progressed through a given chapter linearly instead of randomly throughout a given chapter. This feature was felt to facilitate the student's participation. The key question to ask was whether the linearized (unscrambled) version of the same material would be as effective in helping the student learn the material, as was the scrambled-book version with its emphasis on participation and concentration throughout a given chapter. The student could not half-learn the material in a scrambled-book version; he had to participate to learn each page of instructional material before he could leave the page. The student could skim over the material in the linearized version and thus not learn it thoroughly. The hypothesis was as follows: The students would show superior learning on the scrambled-book version compared to a linearized version of exactly the same instructional material.

METHOD

The programmed instructional material was used as a self-study course with electronic technicians going through the material on their own time at home. There were no classes during work hours, but the technicians were encouraged to ask questions about the same test equipment in the lab at work and to familiarize themselves with its use. Pre- and

[1] A note of thanks is due C. J. Draker of Collins Radio Co., Cedar Rapids, Iowa, for his help in the preparation, training, and data-collection phases of this project. Appreciation also goes to Ken Thompson, United States Naval Training Device Center, for his help in preparing materials.

TABLE 1

LIST AND DESCRIPTION OF VARIABLES

Variable	Level
Fixed	
Sequence of forms in pretest to posttest	2
Book format, linearized or scrambled	2
Card participation, preperforated IBM cards used to record answers, or not	2
Random	
Subjects, electronic technicians	5

posttests were developed and given to measure achievement gain. In addition to developing the linearized version, the scrambled version was revised on the basis of previous tryouts.

The experimental design follows (Winer, 1963). (See Table 1.) Two levels of three independent variables were used in a factorial design with five subjects (*S*s) per cell. The first independent variable was achievement test sequence: this was done to control for a known minor difference in test difficulty. The second independent variable was the format, linear versus scrambled; half of the students got one version, and the balance the other. The next independent variable was student participation, to check on the hypothetical superiority of the scrambled-book version. The participation condition had half of the students using IBM preperforated cards for the recording of answers to the book questions, whether they were using the linearized or scrambled version. The rest of the students had no such cards and thus presumably got themselves less actively involved in the book.

"Subjects" was the last independent variable. The *S*s were all electronic technicians who had been employed for a varying number of years at Collins Radio at Cedar Rapids, Iowa and had different experiential and educational backgrounds. The *S*s were assigned serially to one of the eight treatment conditions as they became available to the study. It was hoped to get six *S*s for each of the eight treatment cells, but several people were not able to complete the book and the posttest satisfactorily. Luckily these people with incomplete records did not occur in the same treatment group so that in the final analysis there were a total of five *S*s for each of the eight treatment cells. Because of a limited number of text books, it was desirable to split the training-testing phase of the experiment into two sections. Accordingly about half of the test *S*s were trained and tested in the summer of 1966 and the second half in the winter of 1967. Data from the two sections were pooled.

RESULTS AND DISCUSSION

The median time for technicians to complete the book on their own time was 12 hr., with extremes ranging from a minimum of 9 to a maximum of 40 hr. The use of card participation did not particularly affect these times. The time involved in pretesting was approximately 1 hr. on company time with another hour likewise in posttesting.

An analysis of variance was made of the gain in test scores from pretesting to posttesting. The posttest scores themselves could have been analyzed but were subject to possible sampling biases in pretest technician ability in the eight treatment cells. This could have been corrected by an analysis of covariance with the two covariables of pretest score and format. The use of gain scores appeared simpler. Consequently, the analysis basically rests on the gains in test scores, as shown in Tables 2 and 3.

Table 2 presents the analysis-of-variance summary for the test gain scores. The basic datum for each *S* was to subtract his pretest score from his posttest score and then to analyze these resultant gain scores. The only significant treatment in Table 2 was that for test sequence. This was expected as the preliminary studies had shown that Form A was harder than Form B. This is reflected in Table 3 where it is seen that a greater gain occurred when Form A was given as pretest and Form B as posttest. This, of course, is to be expected if Form A were harder than Form B. Neither the main effect of the book format nor of card participation was significant. Nor were any of the interactions significant. This lack of significance is also reflected in Table 3 where the averages are given for the treatment conditions. For the difference in book

TABLE 2

ANALYSIS-OF-VARIANCE SUMMARY FOR TEST GAIN SCORES

Treatment and effect	*SS*	*df*	*MS*	*F*
Test sequence (A)	168.10	1	168.10	13.74*
Book format (B)	14.40	1	14.40	1.18
Card participation (C)	0.90	1	0.90	0.07
A × B	8.10	1	8.10	0.66
A × C	14.40	1	14.40	1.18
B × C	4.90	1	4.90	0.40
A × B × C	10.00	1	10.00	0.82
Within	391.60	32	12.24	—
Total	612.40	39		

* $p < .01$.

TABLE 3

SUMMARY OF GAIN SCORES BY TREATMENT

Treatment	Average[a]
Test sequence	
Form A first and B last	5.35
Form B first and A last	1.25
Book format	
Linearized	3.90
Scrambled	2.70
Card participation	
No cards used to record answers	3.15
IBM cards used to record answers	3.50

Note.—Form A = linearized; Form B = scrambled.
[a] N = 20 data per treatment average.

format, linearized versus scrambled, not only was the scrambled-book format not superior to the linearized-book format, as hypothesized, but a little bit of the reverse happened. While insignificant, the linearized format had a slightly greater gain achievement than did the scrambled format. The interpretation can be only a conjecture at this point. The linearized format with its facilitation of review and easy coverage of all answers was just as effective as the scrambled-book format with its somewhat greater participation supposedly. Clearly participation is not part of the answer as the effect of card participation was insignificant although slightly in the expected direction; that is, students using cards to record their answers did have a slightly higher achievement gain than students without the card participation. Or perhaps greater irritation went with the participation in the scrambled book.

Two other studies are relevant at this point. Brown (1966) varied the location of the answer in four different versions of the same programmed instructional package. Typical answer locations were "the next line down" or "on back of the page." He found that the formats with an easy opportunity to cheat resulted in less achievement than where the opportunity to cheat was less frequent or harder. Kapel (1966), on the other hand, found that cheating did not have this detrimental effect. Cheating, defined in terms of opportunity to review material or to search for answers, actually appeared to be superior to where searching was not permitted.

The linearized format used in this study had an easy opportunity to cheat in the sense of being able to search over the following page over all of the answers and pick out just the right one. The person thus could fool himself as well as the experimenter. However, this in itself may have been helpful. A number of students with the scrambled format commented that they looked up the wrong answer sometimes just to find out why it was wrong even though they knew so ahead of time. In other words, even students utilizing the scrambled format learned from looking at the wrong answers or "cheating" in this narrow sense.

In summary the two formats were equally effective in presenting the material, and the reasons for this are not known. Certainly the reason does not appear to involve the supposedly greater participation on the part of the students using the scrambled-book format. Perhaps the answer does lie in the easy availability of all answers in the linearized format for review and optimal search purposes.

There was one methodological flaw in the present study, namely that there was no control group used to evaluate the test-retest effect itself. It is fairly well known that a person will improve on an achievement test if he takes the same test of factual knowledge on a second occasion. This is not the case with esoteric subject matter as has been demonstrated in other studies. The assumption was made in this study that the amount of improvement in knowledge about electronic test equipment would be very slight in the 5- or 6-wk. period between testing unless S

TABLE 4

FORMAT PREFERENCE VERSUS FORMAT STUDIED BY TECHNICIANS

Format studied	Format preference[a]		
	Linear	?	Scrambled
Linear	23	2	1
Scrambled	18	4	3
Totals	41	6	4

[a] $\chi^2 = 4.23$, $p < .05$ by grouping "?" and "Scrambled."

had indeed gone through the textbook. If the study were to be redone, it would be advisable to include a control group to evaluate the test-retest effect over a comparable time period even though it might be hard to obtain *Ss* for this purpose.

Reader preference for one book format versus the other is a final interesting consideration. As students came in to take the posttest, they were asked to fill out a short questionnaire. One of the questions was "Read an item or two in the other book format than the one you had. Which do you prefer, linear or random? Why?" The student tabulations to this question are given in Table 4. Ignoring the question mark response, these results show an overwhelming 41:4 opinion preference in favor of the linear format. The significant chi-square value means that students going through the linear book preferred it more than did students going through the scrambled book. Even though there was no difference in achievement between the two book formats, the tremendous opinion preference for the linearized version has a very practical consideration in terms of student usage. Consequently, the decision was made to go ahead and publish the linear edition (Schuster, 1968).

REFERENCES

BROWN, R. W. Format location of programmed instruction confirmations. *Journal of Programmed Instruction*, 1966, 3(4), 1–4.

CROWDER, N. A. Automatic tutoring by intrinsic programming. In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning*. Washington, D. C.: National Educational Association, 1960. Pp. 286–304.

KAPEL, D. E. An analysis of "cheating" in programmed instruction. *National Society for Programmed Instruction Journal*, 1966, 5(10), 9–12.

SCHUSTER, D. H. *Basic electronic test equipment*. New York: McGraw-Hill, 1968.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1963.

(Received August 28, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 4, 313–318

ABSOLUTE SCALING OF JOB PERFORMANCE[1]

ARTHUR I. SIEGEL, DOUGLAS G. SCHULTZ,[2] M. A. FISCHL, AND RICHARD S. LANTERMAN

Applied Psychological Services, Wayne, Pennsylvania

The logic and development of a technique for evaluating job performance directly in terms of organizational objectives without reference to normative materials are described. The technique involves the establishment of critical performance levels, as derived from a Guttman scale, in terms of stated organizational objectives and relating performance to these levels. The retest reliability, internal consistency, predictability, and concurrent validity of the technique are described.

Throughout the history of psychological and educational measurement, emphasis has been placed on the use of norms for interpreting test scores. It has been commonplace to encounter the statement that a given raw score on a test is meaningless until it is compared to scores made by other people. Textbooks have usually stressed the use of percentiles, standard scores, age norms, etc.

Several recent developments, however, have suggested that there are other, less normatively oriented methods for deriving meaning from test scores. For example, the elaboration of scaling theory has drawn attention to the internal relationships of the items of the score scale itself, somewhat aside from the people from whom the original score data were obtained. Also, systems technology has concentrated on the importance of the functioning of the system as a whole and has evolved concepts of personnel subsystem design and evaluation which relate human performance directly to system performance. In this context, Glaser (1962) has contrasted the use of "criterion-referenced" measures, which tie scores directly to a given criterion behavior, with "norm-referenced" measures, which are interpreted in terms of the score distribution of a representative group of people. With the growing sophistication concerning the problem of assigning meaning to test scores, even current standard textbook treatments of the issue include mention of such matters as types of scales and expectancy tables (e.g., Cronbach, 1960, pp. 69–73).

The purpose of the research reported here was to develop a performance appraisal procedure which would relate each individual's score to a standard, external to the body of personnel, that is, a standard defined by the criterion performance required in order for the particular element of the industrial or military organization to accomplish its organizational objectives.

METHOD

The job constellation under study was the family of enlisted positions associated with naval air electronics (avionics) work, and the method employed consisted of the following steps:

1. Formulation of statements of the specific system objectives of naval air electronics and determination of the importance of their respective contributions to system requirements;
2. Establishment of critical points on a Guttman-type job-performance scale, the critical points representing different levels of satisfying the objectives;
3. Estimation of the reliability of the instrument, derived in Step 2, through fleet administration in six operational squadrons;
4. Investigation of the extent to which the criterion instrument, developed in Steps 2 and 3, is related to various Navy predictor instruments and a job-sample performance test.

Definition and Importance of Objectives

A preliminary list of the objectives of the naval avionics personnel subsystem was prepared following a search of a wide variety of documents concerned with naval and electronic maintenance activities. The search included examination of such sources as bulletins of training course offerings, including prerequisites and purposes of the courses, volumes giving descriptions of the duties of the various occupational specialties, and equipment operation and maintenance manuals. Throughout such literature there

[1] The research reported in this paper was performed under Contract Nonr-4153(00) between the Office of Naval Research and Applied Psychological Services. The cooperation and encouragement of Victor Fields, Sidney Friedman, Charles Hodges, and Martin Wiskoff, Bureau of Naval Personnel, are gratefully acknowledged.

[2] Now at Case Western Reserve University.

were found both direct statements and inferential comments about the avionics personnel subsystem's mission. As a result of this search, four major objectives were drawn up which appeared to encompass the mission of electronics activities carried out in the Navy's air arm. These four objectives concerned the need for *readiness, performance, safety,* and *adaptability*.

This preliminary set of objectives was reviewed with 18 individuals in high-level assignments involving electronics maintenance, training, and supply. Fourteen of the individuals were attached to the Staff of the Commander of Naval Air Forces in the Atlantic Fleet; the remaining four were members of an operational squadron and included the squadron's avionics officer and his maintenance staff. The mean length of the naval service represented in the group of 18 individuals was approximately 19 yr.

Semistructured interviews were conducted with these staff and squadron men at the end of which they were asked to complete the statement: "In operational terms, the electronic maintenance objectives of the Naval Air Fleet are. . . ." On the basis of the information obtained from the interviews and the written statements the preliminary set of objectives was modified, and the final statement of the objectives of naval air electronics was formulated as follows:

1. Operation: To obtain optimum system output when equipment is operated, that is, output characterized by precision and variability appropriate to mission;
2. Performance: To complete any given mission in minimum time at appropriate level of accuracy and reliability;
3. Preparation: To prepare for personnel requirements of present and future equipment, systems, and situations through use of training programs, maintenance of high morale, etc;
4. Readiness: To efficiently maintain self, equipment, and systems in a state of readiness consistent with Fleet requirements;
5. Safety: To carry out duties with maximum protection for men and equipment consistent with mission.

Estimates of the relative importance of these objectives were needed so that they might later be combined in the formulation of overall goals and the resulting establishment of single critical performance points in addition to critical performance points for each objective separately. The objectives were scaled along an importance dimension by the constant-sum method described by Comrey (1950), one of the fractionation methods which produce ratio scale values. The judgments required for the scaling analysis were obtained from 57 commissioned officers and 1 warrant officer, from 30 different operational squadrons, all of whom had current responsibility for avionics activities. Scale values were derived in accordance with Torgerson's (1958, pp. 108–112) modification of Comrey's analytical procedure. This process resulted in importance scale values which ranged between 1.14 and .83. In view of the high degree of similarity among these variables, the five objectives were taken to be equally important.

Establishment of Critical Points on Generalized Job-Performance Scale

Prior research (Schultz & Seigel, 1961) had developed an 8-item hierarchical group of electronic task descriptors which was shown to meet the specifications for a Guttman scale. This generalized job-performance scale of electronics activities had been found to yield a coefficient of reproducibility of .94, and an index of consistency (Green, 1956) of .55. The items of this scale are shown as follows; the scoring values associated with the items are consecutive integers from 1 for "Employing safety precautions on" to 8 for "Employing electronic principles involved in maintenance of."

Items

Employing electronic principles involved in maintenance of . . .
Troubleshooting/isolating malfunction(s) in . . .
Calibrating . . .
Knowing relationship of equipment to other related . . .
Following block diagrams for . . .
Removing . . .
Replacing . . .
Employing safety precautions on . . .

Critical points on this scale were established by adapting the procedures of the psychophysical method of limits. Two threshold points were sought for each objective: one would be the threshold between acceptable and unacceptable performance levels, or the minimum point of *acceptable* performance; the other would be the threshold between acceptable and desirable performance levels, or the minimum point of *desirable* performance. The *S*s for this analysis were 61 noncommissioned officers in 23 operational squadrons who were supervisors in the avionics specialties.[3]

The scale was reproduced on each page of a questionnaire and was headed by one of the five objectives. The *S*s were instructed that the scale items

[3] Aviation electrician's mate, aviation electronics technician, aviation fire control technician, aviation antisubmarine warfare technician, and training devices man.

TABLE 1

POINT JUDGMENTS FOR MINIMUM ACCEPTABLE AND DESIRABLE PERFORMANCE ON EACH NAVAL AVIONICS OBJECTIVE

Job-performance level	Objective										Combined objectives
	Performance		Safety		Preparation		Readiness		Operation		
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	
Desirable											
Descending	6.66	1.20	6.52	1.70	7.11	1.38	7.24	1.14	7.37	1.22	
Ascending	3.50	1.37	3.84	1.62	4.57	1.84	4.47	1.80	5.01	2.09	
M	5.08		5.18		5.84		5.86		6.19		5.63
Minimum acceptable											
Descending	4.37	1.27	4.21	1.74	4.75	1.76	5.16	1.90	5.50	1.61	
Ascending	2.24	1.16	2.63	1.53	3.39	1.79	3.21	1.73	3.75	1.92	
M	3.31		3.42		4.07		4.19		4.63		3.92

Note.—$N = 61$ senior noncommissioned officers in avionics specialties.

were to be interpreted as completing the blank in the sentence: "A technician at this level is capable of _____ most of the equipment typically involved in his rating and is *also* capable of performing all the tasks listed *below* this point on the scale." They were asked to think of journeyman avionics personnel in a typical operational naval air squadron.

When the method of limits is employed to establish threshold points, stimuli are presented in both ascending and descending series, and the final threshold value is taken as the midpoint between the initial threshold values arrived at in the two series. The procedure in this study was to take *S*s, page by page, through the questionnaire. As they considered each objective, they were asked to indicate: (*a*) the place on the scale which described the *lowest* level man they *could make some use* of in meeting the objective, and (*b*) the place which described the *lowest* level man who would be a *desirable* journeyman in the sense that he would *accomplish his assigned tasks so as to fully achieve the objective*. This constituted the ascending series. Then, on additional pages, they were asked to indicate: (*a*) the place on the scale which described the *highest* level man they could *not* make some use of in meeting the objective, and (*b*) the place which described the *highest* level man that they could make some use of but who would *not* qualify as being a *desirable* journeyman in meeting the objective. This constituted the descending series.

Employing the scoring values associated with the items that the *S*s had checked, the mean of the ascending series was computed over all *S*s for the indicated points of minimally acceptable and desirable performance on each objective, as was the mean of the descending series. Then, the mean of these respective means was computed, and its value was taken as the critical point on the job-performance scale. This resulted in two critical scores for each naval avionics objective: the mean of the ascending and descending judgments of the point of minimally acceptable performance became the *minimum acceptable* critical point, and the mean of the two sets of judgments of the minimum point of desirable performances became the *desirable* critical point. These critical points, with the intermediate means and standard deviations upon which they are based, are shown in Table 1.

For each performance level, a mean over all five objectives was also computed and is reported in the column of Table 1 headed "Combined objectives." In this calculation each objective was unit weighted as a result of the determination of essentially equal importance of these objectives reported above. These values, therefore, represent critical points for minimally acceptable and desirable proficiency for the overall objectives of the avionics sector of the Navy.

The overall minimum acceptance judgment was at about Point 4 on the Guttman scale (following block diagrams for), while the overall desirable judgment approached Point 6 (calibrating), a difference greater than most of the standard deviations. At both the minimum acceptable and desirable levels, the operation objective was felt to call for the highest competence and the performance objective the least.

An interesting feature of the data in Table 1 is that the difference between the means resulting from the descending and ascending series is in every case substantial and statistically significant. Either the critical points are not precise points but, rather, more like broad intervals on the basic scale, or else the questions, which were quite different in the two cases, asked of these judges do not reflect merely two approaches to the same point.

Reliability

The job-performance scale, prepared in the format shown in Figure 1, was employed in six operational squadrons of the Atlantic Fleet by immediate line supervisors in evaluating a total of 64 journeyman avionics technicians. Eight weeks later these tech-

TECHNICAL PROFICIENCY CHECKOUT FORM

The man rated on this form should be a striker or petty officer, third class, in one of the following ratings: AE, AT, AQ, TD, AX.

Listed below are eight tasks that a man in one of these ratings may be called upon to perform. You are to indicate whether the man being rated is checked out on each task, i.e., is capable of doing the task "on his own" without direct supervision. Indicate your decision by placing an "X" either in the "CHECKED OUT" or "NOT CHECKED OUT" box to the right of each task description.

TASK DESCRIPTIONS	CHECKED OUT	NOT CHECKED OUT
1. Capable of employing safety precautions on most of this squadron's equipment with which his rating is concerned.	☐	☐
2. Capable of replacing most of this squadron's equipment with which his rating is concerned.	☐	☐
3. Capable of removing most of this squadron's equipment with which his rating is concerned.	☐	☐
4. Capable of following block diagrams for most of this squadron's equipment with which his rating is concerned.	☐	☐
5. Capable of knowing relationship of equipment to other related equipment with which his rating is concerned.	☐	☐
6. Capable of calibrating most of this squadron's equipment with which his rating is concerned.	☐	☐
7. Capable of trouble shooting/isolating malfunction(s) in most of this squadron's equipment with which his rating is concerned.	☐	☐
8. Capable of employing electronic principles involved in maintenance of most of this squadron's equipment with which his rating is concerned.	☐	☐

MAKE CERTAIN THERE IS AN "X" IN A BOX NEXT TO EACH TASK DESCRIPTION

Name of supervisor _____ Rating _____

Name of man rated _____ Rating _____

Squadron _____ Location _____ Date _____

Months supervisor has known man _____

FIG. 1. Appraisal form based upon job-performance scale.

nicians were reevaluated by the same supervisors and the product-moment correlation coefficient was computed between the scores the technicians received on the two evaluations. This correlation coefficient was .97, indicating a satisfactory level of stability for the job-performance measuring instrument over an 8-wk. period.

An instrument which possesses the characteristics of a Guttman scale is internally consistent. However, to determine the effect upon the internal consistency

of this instrument of application to a different sample than the one on which it was originally developed, the data from the sample employed for stability estimation were also subjected to reproducibility and consistency (Green, 1956) analyses. Table 2 presents the results of these analyses, showing the respective reproducibility and consistency values in association with those obtained in the original development of the scale. It is seen that the scalability of the items did not deteriorate upon re-use.

Relationships to Predictors and Performance Test

Copies of the Technical Proficiency Checkout Form and a biographical information form were sent to each of 13 squadrons of the Atlantic Fleet. These were accompanied by a request from the Bureau of Naval Personnel that the Technical Proficiency Checkout Form be completed by immediate line supervisors of the journeyman avionics technicians in the unit and that the biographical information be provided for each man rated. The biographical information consisted of scores on the tests of the Navy's Basic Test Battery, final grades in the particular Basic Specialty Training School attended, and final Basic Specialty Training School class standing. These are the predictor variables commonly employed by the Navy. Class standings were corrected for class size by normalizing to standard scores with a mean of 50 and a standard deviation of 10. Triserial correlation coefficients were then computed between Technical Proficiency Checkout Form scores, trichotomized in accordance with the derived cut points for the combined objectives and each of the predictor variables. Table 3 presents these correlation coefficients. The *N*s vary for the various coefficients because of incomplete data for some cases.

In order to assess the relationship of the Technical Proficiency Checkout Form to another measure of job performance, a job-sample test of knowledge and skill in electronics was constructed. The first part of the test required the technician to identify the safety violations shown in nine pictures. Part 2 consisted of 11 questions concerning seven block diagrams of electrical or electronics equipment. In the third part, the examinee was asked to troubleshoot four malfunctioning electrical circuits in a specially designed piece of apparatus. In the total score on the test, the three parts were weighted equally.

TABLE 2

REPRODUCIBILITY AND CONSISTENCY OF JOB-PERFORMANCE SCALE

Scale	Reproducibility	Index of consistency
Original development[a]	.94	.55
First administration[b]	.97	.57
Second administration[b]	.98	.70

[a] $N = 181$.
[b] $N = 64$.

TABLE 3

TRISERIAL CORRELATION COEFFICIENTS BETWEEN PREDICTOR VARIABLES AND TECHNICAL PROFICIENCY CHECKOUT FORM

Predictor	*N*	*r*
Basic test battery		
General Classification Test (GCT)	471	.14
Arithmetic (ARI)	470	.19
Total, GCT and ARI	470	.19
Mechanical Aptitude (MECH)	470	.22
Clerical Aptitude (CLER)	469	.05
Final grade, specialty training school	408	.15
Final class standing, specialty training school	398	.15

The job-sample test was administered to 50 journeyman naval avionics technicians from 10 operational squadrons, for whom Technical Proficiency Checkout Form appraisals were also obtained. The proportion of these avionics technicians falling in each of the three performance levels, as measured by the Technical Proficiency Checkout Form, was "desirable," 64%; "below desirable but above minimally acceptable," 30%; "not acceptable," 6%. The triserial correlation between the test scores and the Checkout Form scores, as related to the combined objectives, was .40.

These results indicate that the Technical Proficiency Checkout Form is not very predictable from test and training school performance but that it is moderately related to another type of job-performance measure. Such a picture is common in the study of job-performance criteria.

DISCUSSION

This research was designed to expand the interpretive base of job-performance scales to an absolute level by tying them directly to the expressed needs of the Navy. The purpose was to establish the empirical groundwork for converting any appraisal on a checklist form into a meaningful statement regarding the effectiveness of that proficiency level for accomplishing the overall mission of the avionics technician in the Navy. Having established such a relationship, the score of any individual technician could then be interpreted in terms of the expected absolute level of output of that individual, that is, the absolute level of his usefulness to the fleet. The checklist, called the Technical Proficiency Checkout Form, had a satisfactory degree of retest re-

liability, and it retained the Guttman scalability properties across two independent samples (original development administration and two administrations in this research). While the list was not very predictable from basic Navy predictor variables, it was moderately correlated with a job-sample performance test. In terms of the general goals of this research, it seems that the establishment of a reliable, valid, behaviorally based evaluative technique, that is not normatively oriented, has been at least partially achieved. It is also believed that the methods employed are applicable to performance evaluation in other contexts and job families.

REFERENCES

COMREY, A. L. A proposed method for absolute ratio scaling. *Psychometrika*, 1950, **15**, 317–325.

CRONBACH, L. J. *Essentials of psychological testing*. (2nd ed.) New York: Harper & Row, 1960.

GLASER, R. In R. M. Gagné (Ed.), *Psychological principles in system development*. New York: Holt, Rinehart & Winston, 1962.

GREEN, B. F. A method of scalogram analysis using summary statistics. *Psychometrika*, 1956, **21**, 79–88.

SCHULTZ, D. G., & SIEGEL, A. I. Generalized Thurstone and Guttman scales for measuring technical skills in job performance. *Journal of Applied Psychology*, 1961, **45**, 137–142.

TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

(Received August 28, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 4, 319–320

COMPARISON OF OPERATOR PERFORMANCE ON A TRACKING TASK USING DISCRETELY AND CONTINUOUSLY VARIABLE CONTROLS

M. HAMMERTON AND A. H. TICKNER [1]

Medical Research Council, Applied Psychology Research Unit, Cambridge, England

An account is presented of an investigation devised to extend comparison of operator performances using discretely and continuously varying controls to acquisition tasks, and to compare the ease with which Ss mastered the 2 controls. 2 groups, of 12 Ss each, learned 1 control to a stable level, then transferred to the other, which was also practiced to a stable level. It was found that the discretely variable control was harder to master ($p < .05$) and harder to change over to ($p < .01$) when used to the other, although it was confirmed that there was no significant difference between the stable values.

An investigation by Hunt (1966) indicated that there was little to choose between continuously and discretely variable controls in a compensatory task. The present experiment was undertaken to see whether this finding held equally well for acquisition tasks, and to investigate the relative ease with which the two types of control were mastered.

METHOD

Apparatus and task. The apparatus used was the Applied Psychology Research Unit control simulator. The *S*'s task was to move a spot on a cathode ray tube a distance of 5 cm. from an initial rest position to a target, represented by a circle 3 mm. in diameter, and to keep it within that target for 2 sec., continuously. The *S*'s score was the time taken to achieve this.

The *S* exercised control using a thumb joystick: the position of the stick determined the acceleration of the spot. With the continuous control stick, maximum acceleration—given by full deflection either side of center—was .7 cm/sec^{-2}; the sense of the acceleration was the same as the direction of the deflection of the stick from its (lightly spring-loaded) central position; the acceleration varied linearly with deflection. The stepped stick had the same external geometry as the continuous one, but to give an output the stick had to be in one of five spring-loaded indent positions. These positions were at the center of travel, at either extreme, and halfway to either extreme. (The total travel was the same in both sticks.) The output from these step positions was 0 in the center, ±.35 cm/sec^{-2} at the halfway points, and ±.7 cm/sec^{-2} at the extreme points.

Procedure. Two groups of 12 R.N. ratings acted as *S*s for this experiment. The first group was trained for 6 days, having 10 trials per day, using the stepped stick. On the seventh day, they transferred to the continuous stick, with which they were trained for 2 full days, again having 10 trials per day. The second group followed the same procedure with the order of sticks reversed; that is, 6 days with the continuous stick, followed by 2 days with the stepped stick.

RESULTS

For each group of *S*s two measures of stable performance and two measures of transfer were obtained. The stable measures were:

(*a*) The overall group mean of the last five scores before transfer.

(*b*) The overall group mean of the last five scores of all—that is, after transfer.

The transfer measures used (see Hammerton, 1967) were:

(*a*) $a = 100(n - r)/n$, where n is the group mean number of trials needed to reach a stable level with the stick first used, and r is the number of posttransfer trials required to reach a stable performance level with the second stick.

(*b*) $d = 100(1 - T/L)$, where T is the group mean of the first posttransfer score, and L is the group mean of the last pretransfer score.

The quantity n, defined as above, was also used as a measure of speed of initial learning.

It was found for Group 1: stable group mean score with stepped stick = 12.13 sec.; posttransfer stable group mean score with continuous stick = 10.95 sec.; transfer savings

[1] The authors would like to acknowledge the help they received from F. W. Roberts in preparing and maintaining the apparatus.

measure $a_1 = 93\%$; first-shot transfer measure $d_1 = -18\%$; initial learning measure $n_1 = 28$ trials.

It was found for Group 2: stable group mean score with continuous stick = 10.38 sec.; posttransfer stable group mean score with stepped stick = 11.04 sec.; transfer savings measure $a_2 = 80\%$; first-shot transfer measure $d_2 = -226\%$; initial learning measure $n_2 = 20$ trials.

t tests indicate that none of the stable values differs significantly one from another at the .05 level. Applying U tests to the individual a, n, and d values, however, indicates that:

$$a_2 < a_1 \quad (p < .05, \text{ one-tailed})$$
$$d_2 < d_1 \quad (p < .01, \text{ two-tailed})$$
$$n_2 < n_1 \quad (p < .05, \text{ two-tailed})$$

CONCLUSIONS

Hunt's (1966) finding that there is little to choose between the two varieties of control in steady operation is confirmed. It is evident, however, that it is more difficult to master the stepped stick, and more difficult to take it up when used to the other. Consequently, it would seem that the adoption of stepped as opposed to continuous controls is not to be recommended.

REFERENCES

HAMMERTON, M. Measures for the efficiency of simulators as training devices. *Ergonomics*, 1967, 10, 63–65.

HUNT, D. P. Effects of discrete transformations of controller output on human tracking performance. *Journal of Applied Psychology*, 1966, 50, 35–40.

(Received September 5, 1967)

Manuscripts Accepted for Publication in the *Journal of Applied Psychology*

A Set of Basic Interest Scales for the Strong Vocational Interest Blank for Men: David P. Campbell,* Fred H. Borgen, Suzanne H. Eastes, Charles B. Johansson, and Robert A. Peterson: Center for Interest Measurement Research, University of Minnesota, 101 Eddy Hall, Minneapolis, Minnesota 55455.

Emotional Meaning of Four Typographical Variables: Albert J. Kastl and Irvin L. Child*: Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut 06510.

Relative Effectiveness of Programmed Text and Teaching Machine as a Function of Measured Interests: D. D. Cahoon, Lars P. Peterson, and Charles G. Watson*: Program Evaluation Staff, Veterans Administration Hospital, St. Cloud, Minnesota 56301.

A Correlational-Causal Analysis of the Relationship between Expectancy Attitudes and Job Performance: Edward E. Lawler, III*: London Graduate School of Business Studies, 28 Northumberland Avenue, London WC 2, England.

Relationship between Job Satisfaction and Job Involvement: Peter Weissenberg* and Leopold W. Gruenfeld: State University of New York, Vestal Parkway East, Binghamton, New York 13901.

Interaction of Traits and Motivational Factors in the Determination of the Success of Managers: Edwin E. Ghiselli*: Department of Psychology, University of California, Berkeley, California 94720.

Motivation of Research and Development Entrepreneurs: Determinants of Company Success: Herbert A. Waine r and Irwin M. Rubin*: Room 52-538, Massachusetts Institute of Technology, Sloan School of Management, 50 Memorial Drive, Cambridge, Massachusetts 02139.

Task Success, Task Popularity, and Self-Esteem as Influences on Task Liking: Abraham K. Korman*: Department of Psychology, New York University, 21 Washington Place, New York, New York 10003.

Comparative Reliability of Picture Form and Verbal Form Interest Inventories: Richard Blake*: University of Omaha, 60th and Dodge, Omaha, Nebraska 68101.

Complex Vigilance: Relevant and Irrelevant Signals: Irwin L. Goldstein,* William A. Johnston, and William C. Howell: Department of Psychology, University of Maryland, College Park, Maryland 20740.

Consistency and Generalizability of Intraindividual Variability: Ralph F. Berdie*: Student Life Studies, Office of the Dean of Students, University of Minnesota, 2001 Riverside Avenue, Minneapolis, Minnesota 55455.

* Asterisk indicates author for whom the address is supplied.

(Continued on page 330)

Journal of Applied Psychology
1968, Vol. 52, No. 4, 321–324

VEHICLE KINEMATICS AND GAP ACCEPTANCE[1]

PAUL M. HURST, KENNETH PERCHONOK,[2] AND EDMOND L. SEGUIN[3]

Institute For Research, State College, Pennsylvania

Data from traffic photographs under field conditions were analyzed to isolate those variables determining gap acceptability in a merging situation. 28 alternative expressions were computed, involving positional and velocity information for lead, following, and subject vehicles. Regarding each of these indices as a correlate of gap acceptability, correlation ratios were computed between acceptance-rejection behavior and the value of each index at the "decision point." Values of η ranged from .187 to .733. Each of these values, except the lowest, was significant at $p < .01$. In general, the highest correlations were obtained with expressions involving time or speed relationships between subject and following vehicles. Implications are presented for the use of an index and its corresponding η values in the assessment of comparative degrees of "confusion" as a function of roadway and/or traffic-control parameters.

Knowledge of factors influencing driver behavior is necessary for control and prediction of traffic characteristics. In recent years, attention has been given to the effects of interrelationships between vehicles upon the behavior of the drivers concerned. One such setting is the interaction between two vehicles forming a gap and a third which may or may not accept that gap. In this situation, the driver of the potentially merging vehicle may be thought of as judging the relationships between the three vehicles along a single dimension so as to make a decision to accept or reject the gap.

This paper is addressed to the problem of determining those combinations of vehicle kinematics—vehicle separation, relative speeds, and acceleration potentials—which best correlate with observed gap acceptance/rejection behavior. The setting was one in which lane changing was necessitated by blockage of the medial lane on an urban expressway.

In order to determine the predictive power of a given combination of kinematic factors, called gap index (g), reliance was given primarily to η, the correlation ratio. Two qualities desired in a gap index are (a) that different values of the index yield different responses, and (b) that different drivers respond similarly to the same value of the index. The degree to which these criteria are met reflects the degree to which drivers apparently utilize the gap index under study to make their decisions with respect to merging. The value of η is not dependent upon the particular nature of the functional relationship (e.g., linearity), but rather measures the total degree to which the stimulus controls the response. When squared, η measures the proportion of acceptance/rejection behavior accounted for by the gap index.

PROCEDURE

A three-lane section of the John C. Lodge Freeway in Detroit was the site of the data collection. From the roof of a building adjacent to the freeway a fixed observation area, approximately 350 ft. in length, was photographed with a 35 mm. camera at the rate of eight frames per second. Alternate frames were used in the analysis, producing a four per second data rate. During the observation periods a maintenance vehicle had been moved into the downstream end of the observation area, closing the medial lane and thus forcing vehicles in this lane to merge into the adjacent lane. The photographs of resulting vehicle behavior were reduced to usable quantitative form by processing through the Pictorial Data Transducer system developed by the Institute For Research (Hicks, 1964). This system is a film-scanning device coupled with an IBM 024 card punch which enables an operator, by positioning a pair of cross-hairs, to transform photographic information into digital position data in an X–Y coordinate system. The punched-

[1] This investigation was supported by Public Health Service Research Grant No. AC-66, from the Division of Accident Prevention. The authors wish to thank Edward Gervais and Frank DeRose, Jr., and their associates on the John C. Lodge Freeway Traffic Surveillance and Control Research Project, Detroit, without whose cooperation this study could not have been done.

[2] Now at Cornell Aeronautical Laboratories, Buffalo, New York.

[3] Now at HRB-Singer, Inc., State College, Pennsylvania.

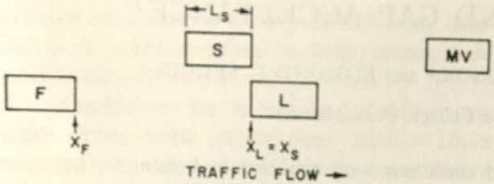


FIG. 1. Diagram of a gap occurrence at the decision point. (S = subject vehicle; F = following vehicle in adjacent lane; L = lead vehicle in adjacent lane; MV = maintenance vehicle blocking medial lane; L_S = length of subject vehicle; X_F = position of following vehicle; X_L = position of lead vehicle; X_S = position of subject vehicle.)

card output is then processed through a battery of special-purpose programs using IBM 1620 and 7074 computers. The computer output consists of point-by-point positional and velocity information for all vehicles in the observation area. For a more complete description of this system, see Hicks (1964).

Measures

From the film segments chosen for study, a total of 107 gap occurrences were analyzed, including 68 acceptances and 39 rejections. This number does not reflect the actual occurrence rate, since to be included in the analysis an "occurrence" had to take place entirely within the observation area. That is, the lead vehicle had to pass the subject vehicle, and the subject vehicle either enter the gap or be passed by the following vehicle, with all these events occurring within the 350-ft. interval. This restriction was necessary in order (*a*) to insure that any gap occurrence labeled as "rejected" was observed over the entire time interval during which acceptance could have occurred, and (*b*) to permit measurement of all appropriate kinematics at a common point, the "decision point." To control motivation, only gaps passing the subject vehicle were considered; gaps which the subject overtook and passed were not counted as occurrences.

The following measures were computed at the "decision point," defined as the instant (to the nearest $\frac{1}{4}$ sec.) that the lead vehicle defining a gap passed a subject vehicle in the closed lane, when the configuration was shown in Figure 1:

MV = maintenance vehicle blocking medial lane.

S = subject vehicle.

L = lead vehicle in adjacent lane.

F = following vehicle in adjacent lane.

X_S = position of subject vehicle (front, longitudinal coordinate).

X_L = position of lead vehicle (rear, longitudinal coordinate).

X_F = position of following vehicle (front, longitudinal coordinate).

L_S = length of subject vehicle.

In addition, the longitudinal component of velocity at the decision point was obtained for each of the three vehicles in question:

V_S = Velocity of subject vehicle.

V_L = Velocity of lead vehicle.

V_F = Velocity of following vehicle.

From the foregoing measures, it was possible to compute the following terms which were utilized in the 27 alternative expressions listed below. Computed decelerations are based upon the assumptions of a reaction time of $\frac{3}{4}$ sec. and a constant frictional coefficient of .6.

$H_F = X_L - X_F - L_S$ = Distance headway anticipated for following vehicle.

$S_F = .75\,V_F + V_F^2/38.4$ = Stopping distance required for following vehicle.

$T_F = .75\,V_F/19.2$ = Stopping time required for following vehicle.

$S_{F\text{-}S} = .75\,V_F + [(V_F^2 - V_S^2)/38.4]$ = Distance required for following vehicle to match velocity with subject vehicle if the latter does not accelerate after the decision point.

$T_{F\text{-}S} = .75 + (V_F - V_S)/19.2$ = Time required for following vehicle to match velocity with subject vehicle if the latter does not accelerate after the decision point.

These terms were utilized in 27 alternative gap indices; a twenty-eighth index, *TH*, the time headway between the subject of following vehicles was also studied. The indices are listed below, with an accompanying value of the correlation ratio, η, expressing the correlation of each index with gap acceptance/rejection behavior. The same 107 gap occurrences form the basis for each calculation.

It should be emphasized that each of these alternative indices, with the exception of *TH*, uses only the information potentially available to the driver at the point in time arbitrarily selected as the "decision point." For example, H_F/V_F is really a measure of projected time headway based on the assumption of constant V_F. Thus, H_F/V_F is the "time headway" projection actually available to the driver at his decision point.

RESULTS

As can be seen from Table 1, the values of η ranged from .187 to .733. The data were analyzed for statistical significance using an arcsine transformation for the proportions of acceptances in each interval. From these data, it is possible to compute a between-cell sum of squares with an approximate chi-square distribution. The transformation was necessary because the dependent variable, acceptance/rejection, is dichotomous.

The $\eta = .187$ value was not significant; each of the other values was significant at $p < .01$.

DISCUSSION

Although some of these indices of gap acceptability have been studied by others (cf.,

TABLE 1

CORRELATION RATIOS FOR VARIOUS INDICES OF g WITH GAP ACCEPTANCE/REJECTION

Expression for g	η
1. V_S	.261
2. V_L	.187
3. V_F	.479
4. H_F	.443
5. S_F	.479
6. T_F	.479
7. S_{F-S}	.473
8. T_{F-S}	.491
9. H_F/V_F	.511
10. $H_F/(V_F - V_S)$	.592
11. H_F/S_F	.537
12. $\{H_F/V_F\}/T_F$	.542
13. $\{H_F/(V_F - V_S)\}/T_F$	.620
14. H_F/S_{F-S}	.585
15. $\{H_F/V_F\}T_{F-S}$	.573
16. $\{H_F/(V_F - V_S)\}/T_{F-S}$	.693
17. $V_S - V_L$	.470
18. $H_F - S_F$	.526
19. $\{H_F/V_F\} - T_F$	.494
20. $\{H_F/(V_F - V_S)\} - T_F$	.632
21. $H_F - S_{F-S}$	.578
22. $\{H_F/V_F\} - T_{F-S}$	.733
23. $\{H_F/(V_F - V_S)\} - T_{F-S}$	.605
24. $\lvert V_S - 1/2(V_F + V_L)\rvert$	.500
25. $(V_S - V_F)/H_F$	.601
26. $H_F/\lvert V_S - 1/2(V_F + V_L)\rvert$	.571
27. $\lvert V_S - 1/2(V_F + V_L)\rvert/H_F$	.533
28. TH	.606

e.g., Glickstein, Findley, & Levy, 1961), direct comparison of results across studies is perilous. The major factor tending to distort such a comparison is the influence, upon η, of the range of dispersion of the stimulus variable. Since η does not specify the functional relationship between stimulus and response values, it is impossible to adjust it for range effects as is often done with the Pearson r. Yet such effects probably do operate in a strong if indeterminate manner.

Thus, the proper comparison of η's is possible only when the same sample of behavior is involved in each computation. In the data from the present study, a fixed sample of occurrences and acceptances, defined by points on photographic film, formed the basis for comparing all of the alternative predictor expressions. Were this sample to be augmented by a number of very large gaps which were generally accepted, and by a number of very small gaps which were generally rejected, the effect would tend to increase the η obtained for any measure having a nonchance relationship with acceptance/rejection behavior. It would still be appropriate, however, to make comparisons within the new sample between η's obtained for various alternative measures of gap size.

Restricting consideration, therefore, to within-sample comparisons, several interesting effects are to be found. First, note that physical gap size or vehicle separation, H_F, was a rather poor predictor of acceptance/rejection ($\eta = .443$, the third from the lowest value obtained). Time headway, TH, was better ($\eta = .606$, the fifth in rank of magnitude). Expressions involving the dynamics of the lead vehicle (L) generally yielded low η's. The four highest values include the terms H_F and V_F. The two highest of these, No. 22 and No. 16, involve a T_{F-S} term which corresponds to time required for F to match velocity with S. In No. 22, T_{F-S} is subtracted from the time headway of F with respect to L, whereas No. 16 represents the ratio between collision course time and T_{F-S}. This collision course time refers to the time required for F to reach S if S were to enter the gap at the decision point, in a position immediately behind that currently occupied by L. A certain strategy on the part of S's driver is suggested here: he is, perhaps, concerned with the ability of F to accommodate him by decelerating before F's headway is exhausted. This could be a consideration regardless of whether S's driver intended to contribute to the velocity matching by accelerating his own vehicle.

Recalling the admonition against direct (as opposed to relative) comparisons across studies, it is nevertheless tempting to compare results with those of Glickstein et al. (1961). Although their data concern on-ramp merging behavior, and are presented in terms of constant-size class intervals (as opposed to the constant-frequency class intervals used here), the comparison may be of heuristic value.

The data from Glickstein et al. are subdivided according to merging by moving vehicles and merging by stopped vehicles; 97% of our data were from moving vehicles. Glickstein et al. used only one measure of gap size, "size of gap presented in seconds," which cor-

responds to our measure of time headway, *TH*. From the data presented by Glickstein et al. in their Table 4, we computed values of η for time headway versus acceptance/rejection, and obtained a value of .463 for moving vehicles and a value of .663 for stopped vehicles. The .463 value is somewhat lower than the value of .606, obtained for *TH* from our primarily moving-vehicle data. The .663 value for stopped vehicles is comparable to some of our better predictive efforts with moving vehicles. A reasonable interpretation would seem to be that the driver of a stopped vehicle is primarily concerned with the time size of the gap as defined by the lead and following vehicles, and his behavior can be fairly well predicted from this datum alone. The driver of a moving vehicle, however, is concerned with the relationship between his speed and certain time or speed characteristics of the gap, so his behavior cannot be well predicted on the basis of time headway alone.

Broad application of any g index must, of course, await cross-validational studies performed under a wide variety of conditions. If it is possible to obtain such an index (or indices) of sufficiently wide generality, we can proceed to compare the degree of stimulus control of gap acceptance across differences in traffic situations: differences in roadway features, traffic control procedures, etc. Suppose, for instance, that Situation A yields a high η between g and gap acceptance versus rejection, and Situation B yields a much lower value of η. This would mean that in Situation A drivers were highly consistent in rejecting gaps of lower g and accepting gaps of higher g. In Situation B, they were less consistent; gaps of lower g were accepted relatively more often, in comparison with gaps of higher g. This would permit the inference that drivers were less confused in Situation A. Consistency, as implied by η, is measurable independently of "threshold" shifts, that is, the tendency to accept or reject more gaps in general. If it is defensible to assume that the driver populations which were sampled in the two situations are similar, then it can be concluded that something about Situation B causes drivers to be more confused than in Situation A, and proceed to derive and test hypotheses as to the source of this confusion. Sometimes the source may be obvious. In other instances it will be more subtle, but regardless of subtlety it should contribute to the understanding and control of driver behavior.

REFERENCES

Glickstein, A., Findley, L. D., & Levy, S. L. Application of computer simulation techniques to interchange design problems. *Highway Research Board Bulletin 291*, 1961, 139–162.

Hicks, W. W., Jr. A pictorial data transducer. Report No. 4, Division of Highway Studies, Institute For Research, State College, Pa., 1964.

(Received September 5, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 4, 325–330

HEURISTIC SIMULATION OF PSYCHOLOGICAL DECISION PROCESSES[1]

ROBERT D. SMITH

Department of Management, Pennsylvania State University

Complex thought processes used by a skilled psychologist dealing with personnel selection and placement are analyzed and programmed in a computer simulation model. Heuristic methods used to limit the total possible set of decision branches in the model are defined and illustrated. Research results indicate a strong relationship between human and machine output with a 94% level of agreement between simulated psychological inferences and human decisions with identical ultimate employment recommendations in 22 of 24 test cases. Implications of research findings on psychological instruction, experimentation, information retrieval, test validation, and general decision making are discussed.

Recent developments in the study of man and machine systems have resulted in significant interdisciplinary achievements. The areas of psychology, industrial management, and operations research have been increasingly integrated through quantitative and scientific methods which aid in the definition, explanation, and prediction of human behavior. One of the most useful of these methods is simulation, a research methodology which facilitates the design and testing of models of physical and human systems. Through the use of simulation, the models can be programmed to yield results which closely approximate real-world phenomena.

Simulation models are of varying types. Some are applied to the study of physical systems such as oil refineries and railroad switching yards. Others deal with man-machine systems such as the Rand Air Defense Simulation in which military personnel and hardware exhibit integrated reactions to simulated changes in the air environment. Yet another type involves human cognition and can be used to replicate the manner in which man makes decisions in solving complicated problems. The purpose of this article is to present the methodology, results, and, most important, the implications of research dealing with this latter form of simulation.

Considerable work in the development of cognitive simulations has been conducted by university and privately sponsored research groups. Colby (1964) at Stanford, Maruyama (1966) at Berkeley, Kleinmuntz (1963) at Carnegie-Mellon, Swenson (1962) at the Mayo Clinic, and Finney (1967) at The University of Kentucky have been instrumental in designing computerized simulations which describe individual psychological characteristics based on some type of test scores.

The present research represents an extension of work previously accomplished. Specifically, this study investigated in depth and subsequently simulated the thought processes of a psychologist involved in the analysis and interpretation of test results, job requirements, and personal characteristics of potential clerical employees. In this type of decision making, the psychologist is faced with ill-structured problems not amenable to solution by algorithmic techniques in which the researcher is able to express an objective function and constraints in well-defined mathematical terms. Instead, it is necessary to rely upon the construction of simulated decision networks which replicate as closely as possible the psychologist's thinking processes.

METHOD

As shown in Figure 1, the experimental model extended the number and types of variables included in previous studies by adding job specifications and personal characteristics of applicants. These additional variables made the task of simulation more difficult since inferences derived from test scores had

[1] This article is based on a paper presented at The College on Management Psychology Session of The Institute of Management Science Fourteenth International Meeting, Mexico City, August 26, 1967.

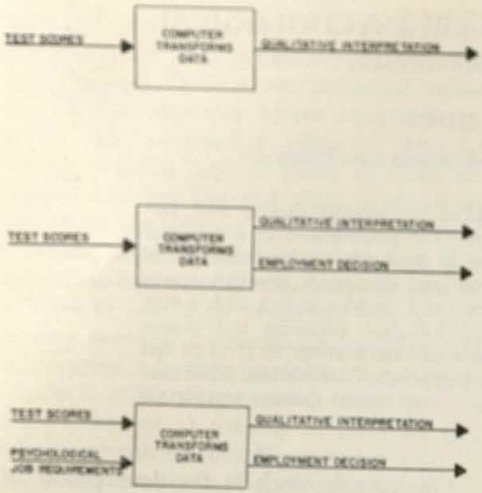


FIG. 1. Evolution of model complexity.

to be viewed in light of specific job requirements and personal characteristics of the individual such as her age, experience, and length of time in her present position.

It was hypothesized that a simulation model could be developed which would replicate the thought processes of a skilled psychologist as he performed the decision task. The Protocol Method was used to gather data (i.e., the analyst's verbalization of thoughts were recorded as he evaluated a series of cases dealing with actual job applicants). The applicants were all female and were applying for various types of clerical positions requiring varying levels of skill. These positions were billing clerk, statistical clerk, clerk typist, receptionist, administrative assistant, and executive secretary. The protocol was tape-recorded, transcribed, and analyzed in depth prior to the development of a computer flow chart defining the decision processes. Finally the model was tested by utilizing a set of cases which were analyzed by the human and processed by the computer.

RESULTS

In all, 24 cases were evaluated by both human and machine methods. In 22 cases the machine selected the same employment recommendation, which was one of the four classifications: hire, reject, hire as a fair risk, or check background further. In one case the machine was not programmed to handle the input data and consequently the human analyst was called in, and, in the other case of nonagreement, the human recommended hiring as a fair risk while the model suggested a further background check of the applicant.

Accuracy of the simulated interpretative statements generated by the computer was judged to be 94% by a skilled analyst not involved in the collection of the protocol. Further, statistical tests of hiring recommendations showed that the probability of obtaining such results by chance was so small that the hypothesis was accepted, and it was concluded that the model did simulate very closely the human analyst's psychological inferences.

Limitations

The preceding paragraphs have very briefly presented the methodology and results of the research.[2] It should be noted that no attempt was made to improve upon the decision-making process of the analyst and it would have been preferable to test a larger sample of cases had they been available. It might be true that a more valid and reliable model would have resulted if the combined decision-making capabilities of a group of analysts had been used. However, rather than spending time in a discussion of the limitations of the study, more significant issues involve the use of heuristics in development of the model and the implications of results on general psychological decision making.

Heuristic Reduction of the Decision Space

In developing the computer model, the matter of relevancy of information became all important. Consider the clerical selection battery used in this experiment. It is comprised of the Otis Mental Ability Test, the Short Employment Tests, the Washburne S–A Inventory, and the Gordon Personality Inventory and Profile. This battery includes 20 different test scores, some of which can range from 1 to 99 such as the Gordon scores while others range from excellent to maladjusted (7-point scale) on the Washburne Test. In all, there are 1,151 possible psychological test scores, in addition to job and personal data, which can be obtained for a single applicant. Thus, it is plain to see that the number of combinations of relationships and elements that can be chosen quickly exceeds manageable proportions. In fact, using the variables

[2] For a complete discussion of research design refer to Smith and Greenlaw (1967).

in this study, it would require 10^{19} outcomes to provide a model encompassing all possible events. This figure approaches the estimated number of atoms in the known universe (10^{18}) representing a programming task of impossible magnitude. To avoid this situation, it is necessary to restrict the boundaries of the system and consider only those relationships which are most relevant to the solution of the decision problem.

To simplify development of the experimental model, the experience of the decision maker was organized into a graphic representation of thought processes. The principles used in the simplification of the model (i.e., to limit the boundaries of relevance) were actually applied by the psychologist and merely transformed into more quantitative terms by the researcher. These principles are often referred to in the literature as "heuristics" taken from the Greek *heveiskein* which means to limit search. Heuristics provide general decision rules or "rules of thumb" which assist researchers in the development of workable models of systems which are within the boundaries of human and machine capabilities. Heuristic solutions to problems are not necessarily the best or optimum solutions since certain alternatives are excluded from consideration which hopefully are irrelevant but, in fact, may not be. Clarkson's (1962) Portfolio Selection and Tonge's (1961) Assembly Line Models were both heuristic in nature as the authors turned their approach from a search for the optimum solution to one that, with high probability, would be suitable for their purposes.

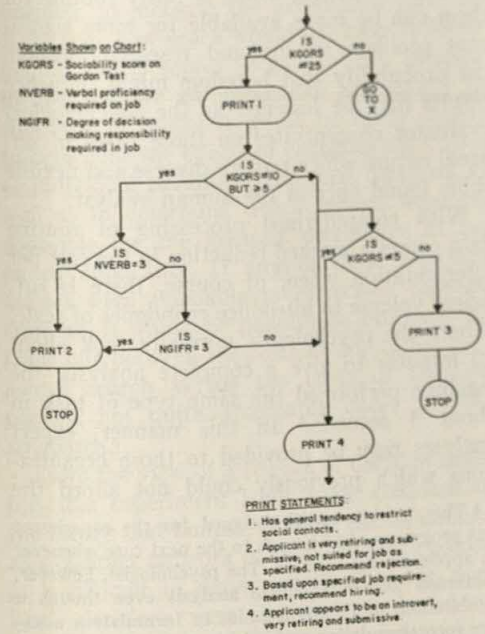


FIG. 2. Partial heuristic decision map for sociability.

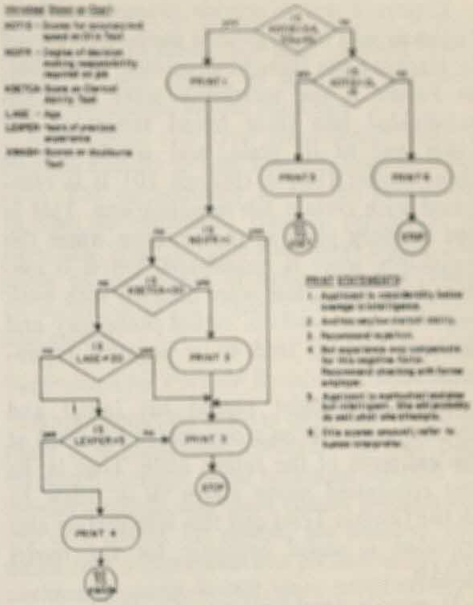


FIG. 3. Partial heuristic decision map for intelligence and ability.

The same heuristic approach was adopted in the present research. Instead of providing for each possible value of a particular variable, relevant ranges for most variables were defined. This type of system definition involved the use of heuristics very similar to those used by the analyst in his interpretations. An example will help to clarify this point. The sociability score on the Gordon Test could range between 1 and 99. The analyst, however, treated only particular ranges for certain job types. For the statistical clerk, a sociability score of 95 or greater coupled with a low cautiousness score led to a negative response since the individual could conceivably be a "social butterfly" and incapable of maintaining the prolonged concentration required for the statistical work.

Figures 2 and 3 provide examples of heuristics as well as the general methodology used in mapping the thought processes. Note that, in Figure 2, the sociability score has been subdivided into three broad ranges at the lower end of its scale, and within one of these ranges (i.e., 6 through 10) it is compared with certain job specifications. This is not the only place in the program where the sociability score is considered, but it is evident here that a low score coupled with work requiring high levels of verbal proficiency and decision making lead to a negative recommendation.[3]

The psychologist placed his initial and greatest emphasis on scores which appeared at the extremes of the rating scale. That is, he first recognized scores above 90 or below 25 on the Gordon Tests and this information also provided a useful heuristic for the model builder.

If certain combinations of variables were relevant but not provided for in the model, the computer was instructed to branch to the next case and leave the exceptional case for human interpretation. In this manner, heuristics were used to reduce the number of combinations of variables within the system to those most relevant for solving the *majority* of cases.

Another type of heuristic was employed which added efficiency to the search process. It was discovered during the analysis of protocol that the psychologist tended to divide his problem into four general components paralleling the four tests used in the battery. He chose mental ability as his first area of investigation and seemed to weigh the Otis scores heavier than other factors for most jobs. Next, he referred to the clerical aptitude results followed by the emotional stability scores, and finally the personality profiles. This does not mean, however, that he followed an unbroken sequence from one test to another. If, for example, he found a girl who performed remarkably well on the Otis Test, he would investigate the Gordon score on "original thinking" as well as the type of job for which she was applying. The intent here was to prevent very intelligent girls from being placed in positions where their ability and interest could lead to poor performance on repetitive or otherwise unchallenging jobs.

The broad approach to sequencing, however, did lead the researcher to design separate computer subroutines for each major area as a starting point in the development of the integrated system. Naturally, there was much subsequent interaction among the subsystems, but the heuristic technique of fractionating the total problem further simplified the work and, more importantly, permitted the simulation to follow a path through the network whereby the most heavily weighted variables were given highest priority. Items which tended to eliminate candidates most often were considered almost immediately, and time was saved since many unacceptable applicants were discovered relatively early in the network.

Implications of Research Results

One of the most important implications of the results of this research is the knowledge that it is possible to program fairly complex psychological decision tasks. This means that time normally spent by highly skilled analysts in decision making which is fairly routine for them can be made available for more significant problem solving and research. Second, the probability that boredom might influence results may be lessened as the psychological evaluator concentrates on those cases of unusual nature which require diverse and flexible skills found only in the human analyst.

With computerized processing of routine data comes a marked reduction in the cost per interpretation when, of course, there is sufficient volume to introduce economies of scale. Where the psychologist normally took about 20 minutes to give a complete analysis, the machine performed the same type of task in about 3 seconds.[4] In this manner, expert analyses may be provided to those organizations which previously could not afford the

[3] Proficiency levels required in specific jobs were expressed on a numerical scale ranging from 1 (lowest level) to 3 (highest level).

[4] The decision model designed for the experiment was programmed to branch to the next case whenever an applicant was rejected. The psychologist, however, continued with a complete analysis even though a combination of factors led him to formulate a negative recommendation somewhere during the course of his evaluation.

services of skilled psychologists. This may be particularly true in personnel offices and guidance centers which sometimes offer submarginal services in the field of psychological analysis.

Simulation models of this type can be useful in the field of computer-assisted instruction. A laboratory approach in basic psychology courses might be established whereby students would verify their beginning and intermediate attempts at analyses against the results of an expert. With the information storage and retrieval capabilities of today's computers, teaching efforts might be made more efficient through the use of feedback data. Professors could evaluate the most common errors made by their students during the laboratory analyses and thereby concentrate upon these areas of weakness in subsequent lectures.

Computerized simulations of psychological decision processes offer other advantages from an information-utilization standpoint. It is now possible to study the effects of policy changes on end results without disrupting the present system. For example, in the personnel model discussed in this paper, one could determine the effect on mental ability standards caused by a labor shortage which required a business to increase the acceptance ratio by 20% (i.e., how much would one have to lower mental ability scores to gain a 20% increase in the number of applicants accepted). In this type of model it would be fairly simple to determine the effects of changing the values of individual parameters on final decisions as well as to determine which factors were responsible for the greatest percentage of rejected applicants. If any of these rejection factors were amenable to correction through training, it would be possible to estimate the trade offs between the costs of additional personnel search versus the costs of training persons not quite acceptable under present standards.

From the view of a science of decision making, this experiment lends some credence to the theory that human thinking can be defined as a complex network of simple binary choices. It is difficult to determine whether the analyst's decisions were or were not based on simultaneous consideration of multivariate criteria. But, by sequencing his thought processes in a series of binary alternatives, it was possible to replicate his final decision even though it was based on relatively complex interrelationships among variables.

Research of this type gives deeper insight into the manner in which people resolve problems. The methodology allows the researcher to map an equivalent thought process at a particular point in time and could permit the study of the effects of aging and experience on decision-making capabilities of individuals and groups.

Suggestions for Further Research

Work should be done to determine whether there is a significant difference in job performance between those personnel selected by human methods and those chosen through the use of a model. Of course, the problem still exists that a representative portion of the experimental group is lost since profit-oriented organizations usually do not hire applicants who have been rejected by skilled interpreters.

The practice of obtaining "objective" evaluations of employee performance still permeates all forms of organizations. Even with the development of modern information systems, managers of all types are faced with the unpleasant task of judging the performance of their workers. Halo and leniency effects predominate. Perhaps it would be possible to design a simulation model whereby the records of each employee could be subjected to a kind of skilled, nonbiased, objective evaluation without fear of political pressure or subjective interpretations.

Results of the study indicate that psychological and other factors can be integrated in a workable decision model. This fact leads one to believe that computer-assisted psychological clinics for low-income classes might be feasible. Admittedly this idea connotes a world of cold impersonalism but there is the possibility that many people would willingly avail themselves of expert services even though these services were not obtained through direct human contact. Further experimentation along these lines would require the development of a model encompassing psychological test results and environmental factors. The model could then be used to screen large numbers of people, some of whom might

exhibit symptoms of present or potential mental illness. A computer-assisted screening center would be used to diagnose cases in early stages for referral to a specialist, thus providing minimum-cost psychological services to broader segments of our society.

REFERENCES

CLARKSON, G. P. *Portfolio selection: A simulation of trust investment.* Englewood Cliffs, N. J.: Prentice-Hall, 1962.

COLBY, K. M., & GILBERT, J. B. Programming a computer model of neurosis. *Journal of Mathematical Psychology,* 1964, 1, 405–417.

FINNEY, J. C. Methodological problems in programmed composition of psychological test reports. *Behavioral Science,* 1967, 12, 142–152.

KLEINMUNTZ, B. Profile analysis revisited: A heuristic approach. *Journal of Counseling Psychology,* 1963, 4, 315–324.

MARUYAMA, M. The use of computers as industrial counselors. *Computers and Automation,* July 1966, 34–39.

SMITH, R. D., & GREENLAW, P. S. Simulation of a psychological decision process in personnel selection. *Management Science,* 1967, 13, 409–419.

SWENSON, W. M. (Chm.), ROME, H., MATAYA, P., MCCARTHY, C., PEARSON, J., KEATING, F., & HATHAWAY, S. Symposium on automation technics in personality assessment. *Proceedings of the Staff Meetings of the Mayo Clinic,* Vol. 37, No. 3, January 1962, 61–82.

TONGE, F. M. *A heuristic program for assembly line balancing.* Englewood Cliffs, N. J.: Prentice-Hall, 1961.

(Received September 12, 1967)

(Continued from page 320)

Skimming Lists of Food Ingredients Printed in Different Sizes: E. C. Poulton*: Medical Research Council, Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.

Knowledge of Score and Goal Level as Determinants of Work Rate: Edwin A. Locke* and Judith F. Bryan: Department of Psychology, University of Maryland, College Park, Maryland 20740.

Organizational Factors and Individual Performance: A Longitudinal Study: George F. Farris*: Massachusetts Institute of Technology, Sloan School of Management, Room 52-590, Cambridge, Massachusetts 02139.

Conceptual and Operational Problems in the Measurement of Various Aspects of Job Satisfaction: Martin G. Evans*: School of Business, University of Toronto, 119 St. George Street, Toronto 5, Canada.

Relation between Birth Order and Being a Beautician: Philip S. Very* and Joseph A. Zannini: Rhode Island College, 600 Mount Pleasant Avenue, Providence, Rhode Island 02908.

Psychological Concomitants and Determinants of Vocational Choice: Kenneth M. Kunert*: Department of Psychology, University of Detroit, Detroit, Michigan 48221.

Ratee Relevance in Peer Nominations: Frank T. Passini* and Warren T. Norman: Department of Psychology, The University of Michigan, Ann Arbor, Michigan 48104.

Punitive Supervision and Productivity: An Experimental Analog: David R. Schmitt: Department of Sociology, University of Washington, Seattle, Washington 98105.

Marginal Productivity Procedure for Staff Selection: Purnell H. Benson*: 123 Milligan Place, South Orange, New Jersey 07079.

Interaction of Achievement Cues and Facilitating Anxiety in the Achievement of Women: W. J. McKeachie*: Department of Psychology, University of Michigan, Ann Arbor, Michigan 48104.

---

* Asterisk indicates author for whom the address is supplied.

Journal of Applied Psychology
1968, Vol. 52, No. 4, 331–334

EFFECT OF PRICING ON PERCEPTION OF PRODUCT QUALITY

J. DOUGLAS McCONNELL [1]

Stanford University

An experiment was undertaken to test whether the price of a product or brand was a determinant of S's perception of the quality of the product. 60 beer-drinking adult Ss made 24 selections from 3 brands of beer over a period of 8 wk. Ss were free to choose whichever brand they wished. Brand names were consonant letters of the alphabet, and the product was identical for all 3 brands. The only real variable was price. Analyses indicated that S's perception of the quality of the beer was significantly affected by the cue supplied, i.e., price.

The theory that price differentials will result in consumers perceiving product-quality differences, where products are in fact homogeneous, has been suggested by Scitovsky (1944–45), Leavitt (1954), and Tull, Boring, and Gonsior (1964). The experiments by Leavitt and Tull, however, were pencil-and-paper tests where *Ss* did not use the brands of products presented in the choice situation. Their findings, therefore, must be treated with reservations. The specific hypothesis tested in the study reported here was that, in the absence of more directly observable cues, consumers accept price as a product characteristic correlating highly with product quality and will alter their perception of a product on the basis of the cues supplied. In other words, it was believed that *Ss* would perceive the higher priced brands as being of higher quality even though the actual product was identical for all brands.

Observation of consumer behavior in the market place provides direct support for this hypothesis. A $5,000 automobile is perceived as a better quality car than one costing $3,500; a $3.00 pair of stockings is perceived as being superior in quality to an $.85 pair, yet both may "run" within 5 min. of being worn. An $80 gown is superior to a $45 gown. The phrase "you get what you pay for" is accepted as a truism in the American culture and those of most highly industrialized nations.

In view of this, the hypothesis appeared to be soundly based. It was believed that if a "price-quality" cognition existed the *Ss*, being unable to change the "price-brand" stimulus, would change their perception of the product to maintain consonance between their cognitions and perception.

METHOD

Subjects

Sixty adult married students, or wives of students, who classified themselves as beer drinkers, were chosen at random from Stanford University Married Student Housing and served as *Ss*. Median age was 27 yr. with a range of 21–55 yr. Family income ranged from $2,500 to $15,000, with a median value of $5,000. Forty-four *Ss* were male and 16 female.

Product

The product used in the experiment was beer which was obtained from a California brewer with the approval of the California Department of Alcoholic Beverage Control and the United States Internal Revenue Service.

An order was placed with one brewer for 60 cases of lager-type beer, bottled in unmarked, nonreturnable "glass cans," and sealed with a plain seal. It was specified that all beer come from the same batch number so that the product would be physically as homogeneous as possible.

It was decided to restrict *Ss*' range of choice in the experiment to three brands of beer. To identify these three brands, cases of beer were drawn from inventory as they were required and labeled at random with white rectangular self-adhesive labels lettered L, M, or P in red-felt marker ink. The letters, following Tucker (1964), were selected for ease of identification and memory and because each is a consonant falling in the middle of the alphabet and occurs with approximately the same frequency in usage of the English language. Nonsense syllables, such as suggested by Davidson, Suppes, and Siegel

[1] Now at Stanford Research Institute, Menlo Park, California. The article is based on research undertaken as part of the author's unpublished doctoral dissertation, Stanford University, May 1967.

TABLE 1

DATA FOR CONTINGENCY TABLE ANALYSIS

Ratings from Likert-type scale	Price levels of beer		
	High	Medium	Low
Undrinkable 0	4	1	4
Poor 1	8	21	20
Fair 2	26	22	23
Good 3	15	12	9
Very pleasant 4	7	4	4
Total score	133	117	109
Mean score	2.23	1.93	1.80

(1957), were considered but rejected on the grounds that if they were "association free" they would be difficult to learn—an undesirable attribute for any brand name. It was decided to risk the possibility that the letters chosen might have meanings for some *S*s to facilitate learning of the brand names.

The price was shown to the customer in two ways. One, a card behind each test bottle told the price of that brand per six-pack since most consumers see it in the stores that way. As *S*s were not paying for the bottle of beer they selected at each trial, the situation was considered to be very artificial and not comparable to the purchase decision situation that usually confronts a shopper. To minimize the effect of these free or "windfall" goods in the selection process, the less expensive brands had to represent a financial saving compared with the high priced brand, equivalent to the six-pack price differentials. This was achieved by taping the per bottle price differential of 5¢ on the low priced brand and 2¢ on the medium priced brand. Thus, an *S* selecting the low priced brand would be in a position similar to a consumer who purchases a low priced brand and has more change in his pocket than he would have if he bought a high priced brand.

Procedure

The *S*s were called on in their own apartments at prearranged times. A number were called on between 8:30 A.M. and 9:15 A.M., while the greater majority were visited between 5:00 P.M. and 7:30 P.M.

On each visit *S*s selected one bottle from the three brands (L, M, and P) which were set out on a wooden tray and backed by a card stating the price per six-pack. For each trial the positions on the tray were changed so that any selections based on position on the tray would be completely randomized. No brand occupied the same position on the tray for any two successive trials, and each brand occupied each position with equal frequency over the 24 trials. The 24 trials were conducted at the rate of three per week for 8 wk. The *S*s undertook to drink each bottle selected at any time convenient to them before making their next selection. After their final selection in Trial 24, *S*s were presented with a questionnaire form to be completed. The *S*s were asked on this form to rate each of the three brands of beer on a 5-point Likert-type scale. They were also asked to select from a panel of words, commonly used to describe beer, the three words that best described each brand. The words in the panel are shown in Table 3 and have generally favorable and unfavorable connotations with respect to beer.

RESULTS

The data were analyzed in three ways. A simple contingency table analysis (Mood & Graybill, 1963) was made to test the relationship between the ratings obtained and the three price levels of the brands of beer. The

TABLE 2

SUMMARY OF RATING SCORES BY BRAND AND VALUES FOR *t* USING METHOD FOR CORRELATED OBSERVATIONS

Numerical value per rating point	Brand M ($1.30)		Brand L ($1.20)		Brand P ($0.99)	
	No. responses	Score value	No. responses	Score value	No. responses	Score value
0	4	0	1	0	4	0
1	8	8	21	21	20	20
2	26	12	22	44	23	46
3	15	45	12	36	9	27
4	7	28	4	16	4	16
Total score		133		117		109
Mean score		2.23		1.93		1.80

Note.—According to the Student's *t*, mean differences between Brand M versus Brand P = 2.46, $p < .01$; Brand M versus Brand L = 1.64, $p < .06$; Brand L versus Brand P = .93, *ns*.

TABLE 3

PANEL OF WORDS USED BY SUBJECTS TO DESCRIBE THE THREE BRANDS OF BEER

Favorable (implied high quality)	Unfavorable (implied low quality)
Tangy	Flat
Rich flavored	Biting
Smooth	Acidy
Full bodied	Watery
Light	Bitter
Dry	Sour

data for this analysis form Table 1. The ratings of "undrinkable," "poor," "fair," "good," and "very pleasant" were assumed to be monotonically increasing and were given values of 0, 1, 2, 3, and 4.

The contingency table analysis produced highly significant findings, the χ^2 distribution value being significant at $p < .005$ level, thereby indicating the existence of a price-quality relationship.

The second test took the mean ratings obtained for each of the three price levels and tested each pair of means for significant differences, using the Student's t distribution and method for correlated observations (Winer, 1962). This method was used because the ratings for each brand could not be assumed to be independent as each S rated each brand. The results of this analysis are given in Table 2. A significant quality difference was perceived between the high and low priced brands. The interesting aspect of this analysis, however, was that the middle priced brand was not perceived as being of significantly higher quality than the low priced brand. The "perceptual distance" between the high and middle priced brands was greater than between the middle and low priced brands, although the actual price differential would have led one to suspect that the reverse should have been the case.

TABLE 4

BASIS OF DETERMINING NUMERICAL SCORE FROM WORDS CHOSEN BY SUBJECTS TO DESCRIBE BRANDS

No. responses		Numerical score
Favorable words	Unfavorable words	
3	0	8
2	0	7
2	1	6
1	0	5
1	1	4
0	1	3
1	2	2
0	2	1
0	3	0

TABLE 5

SUMMARY OF ANALYSIS OF VARIANCE FOR FAVORABLE AND UNFAVORABLE WORDS ATTRIBUTED TO BRANDS

Source of variation	*SS*	*df*	*MS*	*F*
Total	1800.20	179		
Between brands				
M, L, P	64.53	2	32.265	3.29*
Within	1735.65	177	9.801	

* $p = .05$, $F = 3.07$, $df = 2/120$.

The third test of the price-quality relationship used *S*s' selections from the panel of words with favorable and unfavorable connotations listed in Table 3.

To examine *S*s' selections statistically, a numerical score was assigned to each *S*'s selection on the basis of the number of favorable and unfavorable words chosen for each brand. This is shown in Table 4.

This test permitted *S*s to be regarded as "blocks," and for price to be regarded as "treatments," and a one-way analysis of variance (Winer, 1962) to be carried out. The results were significant at the $p < .05$ level as indicated in Table 5.

Apart from this analysis of *S*s' word choices, the overall direction of selections is readily ap-

TABLE 6

RELATIONSHIP BETWEEN PRICE OF BEER AND DESCRIPTIVE WORDS

Type of words chosen	Brand of beer and price per six-pack		
	M ($1.30)	L ($1.20)	P ($0.99)
Favorable	93	73	57
Unfavorable	71	82	101

Note.—Totals chosen are number of words.

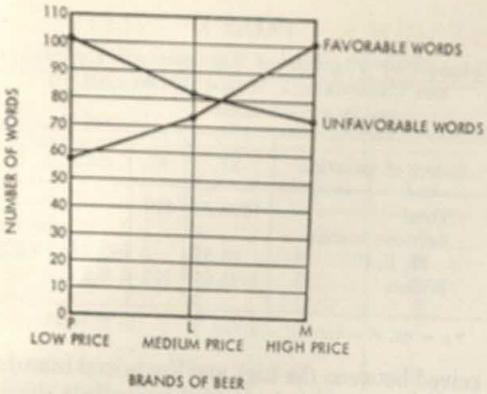


FIG. 1. Plotting of favorable and unfavorable descriptive word responses by brand and price.

parent from a cursory examination of Table 6 and Figure 1.

DISCUSSION

Price was used by *S*s as a cue to product quality. With a physically homogeneous product and unknown brand names (which had so little meaning that many *S*s never used them), the highest priced brand was perceived as being a "better quality" product than the other two brands. Brand L, the middle priced brand, was perceived as being marginally better than the lowest priced brand. In fact, if perception and reality had correlated perfectly, each *S* would have tried each brand once and then chosen Brand P (the lowest in price) for the remaining trials. In this way *S*s would have maximized their utility, receiving the same beer and 5¢ at each trial instead of something less. The *S*s, however, selected Brand M (high price) for over 41% of all trials and Brand L (middle price) for 25% of all trials. Price, in the absence of other cues, was a powerful determinant of how the brands were perceived.

Perceived quality did not vary linearly with price. The greater price disparity existed between Brand L and Brand P, a total difference of 3¢ a bottle. The price difference between Brand M and Brand L was only 2¢ a bottle. However, the perception of quality was such that Brand L and Brand P were perceived as much closer than Brand L and Brand M. This study found no causal explanation for this phenomenon. A first-order Markov analysis of brand switching revealed that Brand L lost to both Brands M and P with the greater movement to Brand M. This analysis tended to confirm the significance of the ratings obtained for quality.

The study strongly supports the earlier work that price is an independent variable on which perception of a brand or product depends. The extent to which this may be generalized beyond the United States culture and over a wide range of products is unknown. Further research is clearly indicated.

REFERENCES

DAVIDSON, D., SUPPES, P., & SIEGEL, S. *Decision-making: An experimental approach.* Stanford: Stanford University Press, 1957.

LEAVITT, H. J. A note on some experimental findings about the meaning of price. *Journal of Business,* 1954, **28**, 205–210.

MOOD, A. M., & GRAYBILL, F. E. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1963.

SCITOVSKY, T. Some consequences of the habit of judging quality by price. *The Review of Economic Studies,* 1944–45, 12(2), 100–105.

TUCKER, W. T. The development of brand loyalty. *Journal of Marketing Research,* 1964, **1**, 32–35.

TULL, D. S., BORING, R. A., & GONSIOR, M. H. A note on the relationship of price and imputed quality. *Journal of Business,* 1964, **37**, 186–191.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received September 15, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 4, 335–340

COMPARISON OF QUESTIONNAIRE RESPONSES WITH AND WITHOUT PRECEDING INTERVIEWS

CLAYTON P. ALDERFER [1]

Graduate School of Business and Public Administration, Cornell University

This study reports the results of a field experiment carried out in an industrial organization designed to test the impact on fixed-alternative questionnaire answers of having some respondents be interviewed by a series of open-ended questions covering the same topics prior to taking the questionnaire. Support was provided for 1 of the 2 hypotheses tested. More negative data were provided about peer relations by those who were interviewed before doing the questionnaire than by those who were not. This finding held for both low- and high-skill employees. There was no difference in the data about superiors under the 2 methodological conditions. The results were interpreted in the context of other research in this area, and 2 continua for explaining field method effects on questionnaire validity were proposed.

The place of the interview in social science research has been a subject of some controversy. Some investigators have cautioned against the use of the technique because of the possibility that an interviewer might be able to lead a respondent to bias his answers toward those the researcher wanted (Dunnette, 1962). Others have criticized the interview on the ground that it provides data only on a respondent's perception of events (Becker & Greer, 1957). Nevertheless, a fairly extensive literature has developed around styles and types of interview (Kahn & Cannell, 1957; Richardson, Dohrenwend, & Klein, 1965; Whyte, 1960). Another methodological debate and inquiry has concerned the use and misuse of questionnaires. Public criticism has been voiced against "survey sickness," the agonizing pursuit of questionnaire data to answer trivial questions in the misguided quest of science (Hechinger, 1966). There have also been a number of studies directed toward empirically determining some of the conditions which facilitate obtaining valid data by the use of questionnaires.

One variable studied has been the effect of anonymity of respondents. Although a number of studies have been reported, the results are not entirely consistent. Pelz (1959), for example, compared the questionnaire responses of *S*s who were completely anonymous to those who were identified but confidential. The basic conclusion was that there were no differences between the two methods of administering questionnaires in terms of the attitude scores obtained. The investigator, however, cautioned that these results would probably have been different if the study were done under different conditions, for example, if careful preparation had not been carried out, if there were fear of layoff, or if there were widespread hostility toward management. Klein, Maher, and Dunnington (1967), on the other hand, found that questionnaire responses were more positive for identified than for anonymous respondents. The answers were also more positive when a plant manager told the identified *S*s to use an identifying code than when a member of the corporation personnel research group made the same request. Specific questions also had a greater tendency to be positively distorted, namely, ones about pay and evaluation by one's superior.

The Klein, Maher, and Dunnington study was carried out entirely by company staff. Dunnette and Heneman (1956) compared the responses of employees to whom a questionnaire was administered by their personnel manager with those to whom the same questionnaire was administered by an independent university researcher. They found significant positive distortion on general morale items and on supervisory-subordinate issues by the respondents to whom the questionnaire was administered by the insider.

[1] The writer would like to thank Richard A. Dunnington and Chris Argyris, Fritz Steele, and William F. Whyte for their comments on an earlier version of this paper.

Still another way to look at these methodological issues is to ask about the effects of simultaneously using more than one method to investigate a problem area. For example, do respondents answer a fixed-alternative questionnaire differently if they have been interviewed prior to taking the questionnaire? The present study reports a field experiment designed to answer that question.

Theoretical Issues

One point of view about the research process suggests that it contains elements of mutuality (Argyris, in press; Maslow, 1966). If a researcher wants valid data, it helps if he is perceived as a trustworthy person. If he wants complete data, it helps if he makes himself accessible to observation, too. Of course, this is not the only way to conceive of the research process. One might also suggest that respondents do not want to know the researcher and prefer that he remain distant, uninvolved, and presumably neutral. How might one test between these two points of view using interview and questionnaire methodology?

Suppose the researcher conducted an open-ended interview with each respondent before asking him to complete a questionnaire. The interview might serve a number of functions: (*a*) the respondent would become familiar with the topics of the study, (*b*) he would be able to answer the questions in his own terms rather than be forced into fixed-alternative answers on the questionnaire, (*c*) he would be able to ask questions of the researcher without being observed by other members of the organization, and (*d*) he would have considerable time to "size up" the investigator and to find ways to test his trustworthiness.

If an interview provided a respondent with a mechanism whereby he could trust the researcher more, how would this affect questionnaire responses? The impact of trust generally is to provide a climate of psychological safety, thereby allowing a person to talk more freely about potentially threatening issues. Some previous research provided a basis for a priori predictions about which research topics would be more likely to be distorted if adequate researcher-respondent trust were not present. Argyris (1962) has found that interpersonal issues, such as dealing with emotions and facing conflict, tend to be difficult for many organization members in their day-to-day activities. If these issues present problems for people in their regular dealings with each other, then one would suggest that they would also be quite threatening in single encounters with an outsider whose trustworthiness might be open to question. In addition to the general threat posed by interpersonal issues, there is also the fact that some of the major issues in the client-researcher relationship are questions of trust, confidence, and respect. So one would expect the research encounter itself to evoke concerns about interpersonal issues. Consequently, the attitudes which would be expected to be more fully recounted if additional trust could be developed in the researcher-client relationship would be those dealing with interpersonal relationships. If the attitudes are more fully described, their overall sentiment would be more negative because the factors people would withhold due to lack of trust would be the negative aspects. The following hypotheses, therefore, were subject to test in a field experiment.

Hypothesis 1: Respondents who were interviewed by a series of open-ended questions covering the same subjects as those of the questionnaire would give more negative accounts of satisfaction with respect from superiors than respondents who were not interviewed before taking the questionnaire.

Hypothesis 2: Respondents who were interviewed by a series of open-ended questions covering the same subjects as those of the questionnaire would give more negative accounts of satisfaction with respect from peers than respondents who were not interviewed before taking the questionnaire.

METHOD

Procedure

The study was carried out over a period of 3 mo. in a manufacturing organization employing approximately 1,700 people. Three shifts were in operation, and data were gathered on each of the shifts. A week preceding the time when an *S* was asked to appear to participate a notice was sent to him through his first-line supervisor. The memorandum indicated the time and place for the person to appear and indicated whether he would be interviewed or asked to complete a questionnaire.

When a person came for an interview he was greeted by the investigator and informed about the purposes and topics of the study. While this explanation proceeded, a small tape recorder was setting on a nearby table. The case was open so that it could be clearly seen that it was not turned on. When the initial introduction had been completed, the interviewer asked for the respondent's permission to tape the interview. The person was assured that no one in the company would ever be allowed to listen to the tapes.

All of the interviews were conducted by the writer. Almost without exception the questions were asked in the same order. Some sample questions were, "What do you like to get out of a job?" or "Would you describe your boss for me?" There were 15 standard questions. Additional probing questions were used to clarify what a respondent meant when it seemed unclear to the investigator.[2]

Questionnaires, for people not interviewed, were distributed during meetings attended by 10–20 persons at one time. The meetings were held between shifts and were attended by people who were leaving after having completed their work for the day. No more than three people from a given department were asked to attend a questionnaire administration at one time. No more than two people from a given shift in a given department were interviewed in 1 day.

Persons who were interviewed were given a questionnaire and a stamped envelope addressed to the investigator after the interview had been completed. The questionnaire had a code number on the back. The *S* was told that this number existed so the researcher could keep all the information provided by each person together. The confidentiality of the study was reiterated at this point.

Persons who attended questionnaire meetings either completed the instruments in the meetings or, if there was not sufficient time, also received a stamped envelope addressed to the researcher. Questionnaires which were completed in the meetings had no identifying labels; neither names nor code numbers were provided. Where the questionnaires were to be mailed back, then identifying numbers were used. The *S*s were informed that the investigator wanted to be able to remind them in case they forgot to send back their questionnaires. Confidentiality was reaffirmed. Approximately 60% of the questionnaires were completed in the meetings.

Subjects

The respondents in this study were all nonsupervisory employees. Two groups of *S*s are distinguished according to skill level. People holding the low-skill jobs had duties which included: custodial work, highly repetitive assembly-line operations, machine tending in continuous process operations, inspection, and stock accounting. The high-skill group included: the trade groups such as mechanics, machinists, and electricians; quality control technicians who developed new inspection standards and worked away from the production lines; and advanced machine operators who tended highly complex machines and carried out selected staff functions as well.

The *S*s were randomly selected from the list of employees in the organization. Approximately 90% of the people who were selected participated in the study. Roughly one-third of each skill group were randomly assigned to be interviewed before taking the questionnaire. Forty-four low-skill and 39 high-skill employees were interviewed, while 88 low-skill and 75 high-skill took only the questionnaire.

Dependent Variables

The data used to test the experimental hypotheses were taken from the questionnaires. Sample items from the four scales are presented below. They appeared in a Likert scale format. Agree or disagree after the item indicates the direction scored for satisfaction. A full listing of the items in each scale is given elsewhere; the total number of items in each scale is indicated (Alderfer, 1967).

Pay satisfaction: (7 items).

Compared to the rates for similar work here, my pay is good. (agree)

I find it hard to do some of the things I enjoy most on the pay I receive. (disagree)

Respect by superiors satisfaction: (8 items).

My boss takes account of my wishes and desires. (agree)

My boss will play one person against another. (disagree)

Respect by peers satisfaction: (8 items).

I can count on my co-workers to give me a hand when I need it. (agree)

My co-workers are uncooperative unless it's to their advantage. (disagree)

Use skills and abilities satisfaction: (12 items).

My job requires that a person use a wide range of abilities. (agree)

In my job I have same thing to do over and over. (disagree)

RESULTS

Table 1 presents the means and standard deviations for the two skill groups under the different experimental conditions on each of the four attitude scales. The data do not provide support for Hypothesis 1, for either skill group. There are no statistically significant differences on the questionnaire measure of respect by superiors satisfaction between the two experimental conditions in either skill group.

There is support for Hypothesis 2, however. The low-skill employees who were in-

[2] A complete account of the interview questions and introductory procedures is given elsewhere (Alderfer, 1966).

TABLE 1

QUESTIONNAIRE MEASURES OF NEED SATISFACTION FOR QUESTIONNAIRE-ONLY AND INTERVIEW-AND-QUESTIONNAIRE GROUPS AT TWO SKILL LEVELS

Group	N	$\bar{X}$	SD	Z
Pay satisfaction				
Low-skill questionnaire-only	88	−.03	1.01	−.65
Low-skill interview-and-questionnaire	44	−.16	1.13	
High-skill questionnaire-only	75	.12	.97	.72
High-skill interview-and-questionnaire	39	.25	.91	
Respect by superiors satisfaction				
Low-skill questionnaire-only	88	.25	.79	.72
Low-skill interview-and-questionnaire	44	.38	.90	
High-skill questionnaire-only	75	−.17	1.20	−.70
High-skill interview-and-questionnaire	39	−.31	1.10	
Respect by peers satisfaction				
Low-skill questionnaire-only	88	.10	1.05	−1.95*
Low-skill interview-and-questionnaire	44	−.33	1.28	
High-skill questionnaire-only	75	.23	.74	−2.74**
High-skill interview-and-questionnaire	39	−.29	1.06	
Use skills and abilities satisfaction				
Low-skill questionnaire-only	88	−.43	1.08	−.11
Low-skill interview-and-questionnaire	44	−.45	.93	
High-skill questionnaire-only	75	.20	.80	−.50
High-skill interview-and-questionnaire	39	.12	.87	

Note.—The attitude scale scores were standardized with a mean of 0 and standard deviation of 1 for the whole organization. Negative signs on the means refer to the dissatisfaction end of the scale; positive signs to the satisfaction end.
* $p < .026$, one-tailed test.
** $p < .005$, one-tailed test.

terviewed prior to completing a questionnaire gave significantly less satisfied accounts of respect by peers than the low-skill employees who were not interviewed. A similar finding holds for the high-skill group. The high-skill respondents who were interviewed prior to taking a questionnaire showed significantly more negative respect by peers satisfaction attitudes than the high-skilled employees who were not interviewed.

Data on the other two attitude scales, pay satisfaction and use skills and abilities satisfaction, show no significant differences for either job group as a function of the experimental conditions.

Perhaps it should be mentioned that the four dependent variables are essentially uncorrelated with each other. Their median correlation is zero; the extreme values are .02 and −.04. Therefore, tests on the separate scales are meaningful, and positive or negative findings on one scale are essentially independent of results on other scales.

Also, the reliabilities of the four scales are reasonably high. Spearman-Brown estimates for the pay, superiors, peers, and abilities scales are, respectively, .88, .84, .79, and .88. Therefore, it is unlikely that failure to detect differences between experimental conditions can be attributed to lack of reliability in the dependent measures.

DISCUSSION

The experimental findings provide support for only one of the hypotheses. This support, however, was reasonably strong in that it held for both the low- and high-skill groups. Why was there not support for the respect by superiors satisfaction variable as well as for the respect by peers satisfaction measure? Two explanations are suggested.

The first implies that the experimental conditions were not strong enough. That is, while having an interview prior to doing a questionnaire raised the respondents' trust enough so they revealed more negative attitudes about their co-workers, the procedure was not enough to have a similar effect on attitudes about superior-subordinate relations. This explanation assumes that discussion of superior-subordinate relations is more threatening than discussion of peer relationships and therefore requires a higher level of researcher-client confidence for the respondents to feel free to admit to negative attitudes about their bosses.

The second explanation says that as far as superior-subordinate relationships were concerned, the additional trust made no difference because S was not very threatened. This is to imply that the respondents said all there was to say on the questionnaires, and the increment of trust which interviews might have added had no payoff in more valid data about superior-subordinate relations. Following the Dunnette and Heneman (1956) findings, one might reason that the investigator being an outsider may have been all that was necessary to obtain complete data on this topic. In terms of the data of this study, this explanation assumes that discussion of superior-subordinate relations was *less* threatening than discussion of peer relationships.

If the second explanation is valid, then combining the data from this study with those obtained by Dunnette and Heneman (1956) and by Klein, Maher, and Dunnington (1967) suggests two generalizations about the impact of field-research methodology on the validity of questionnaire responses.

It should be emphasized, however, that these generalizations are intended to be quite tentative. They represent one way of bringing some order out of the results of the present and related studies on this subject. Since they are not empirically confirmed, they may best be seen as hypotheses suggestive of profitable directions for future research.

The first generalization concerns a continuum of research methods ranging from those which seem to lead to more response distortion to those which lead to less bias on questionnaires. The following order from most to least distortion is suggested:

(*a*) Identified questionnaires with administration in part by line management;

(*b*) Unidentified questionnaires with administration in part by line management;

(*c*) Identified questionnaires with administration by internal (personnel) staff;

(*d*) Unidentified questionnaires with administration by internal (personnel) staff;

(*e*) Identified questionnaires with administration by external (university or research institute) researchers;

(*f*) Unidentified questionnaires with administration by external (university or research institute) researchers; and

(*g*) Respondents interviewed by external researcher prior to filling out questionnaires.

The reader will note that although this order is presented as a single dimension, more than one variable enters the definition of the points. For example, who the administrator is receives a heavier weight than whether the respondent is identified. Such an assumption involves going a bit beyond what the results of the studies cited above and the present results strictly show, but the extension does not seem unreasonable. Furthermore, there is no way of knowing what effect interviewing preceding the questionnaire administration would have if the interviewer were an internal researcher.

The second tentative pattern emerging from combining this study results with those of other writers is an ordering of issues whose validity increases as one moves along the methodological continuum. As the method provides more respondent-researcher trust, the issues seem to have more valid data provided about them in the following order:

(*a*) Pay satisfaction;

(*b*) Superior-subordinate relations, particularly around subjects of mutual trust and equity; and

(*c*) Peer relations, concerning such issues as acceptance, respect, and trust.

It would seem that there is a basic dimension underlying the ordering of these issues. As one moves from pay, through superiors, to peers, the object of interest changes from a material factor external to an important person who (at least formally) is somewhat distant from the respondent to co-workers who, quite likely, are closer to the respondent. The underlying continuum appears to be intimacy. It seems that negative data about more intimate issues by questionnaire become more available as a safer and closer relationship is developed between researcher and respondent.

When satisfaction with pay and satisfaction with use of skills and abilities did not show differences between the experimental conditions, indirect support was provided for the idea that the preceding interview affects primarily questionnaire responses about interpersonal relationships. From these data, there is little reason to suggest that an external researcher should invest the time and money required to interview people if he does not want to obtain data about interpersonal relationships in the organization. On the other hand, where data about interpersonal relationships are desired, there is reason to believe that preceding questionnaire administration with a semistructured interview can reduce distortion of questionnaire responses.

REFERENCES

ALDERFER, C. P. Convergent and discriminant validation of satisfaction and desire measures by interview and questionnaire. *Journal of Applied Psychology*, 1967, **51**, 509–520.

ALDERFER, C. P. Differential importance of human needs as a function of satisfactions obtained in the organization. Unpublished doctoral dissertation, Yale University, 1966.

ARGYRIS, C. Some unintended consequences of rigorous research. *Psychological Bulletin*, 1968, in press.

ARGYRIS, C. *Interpersonal competence and organizational effectiveness*. Homewood, Ill.: Dorsey Press, 1962.

BECKER, H. S., & GREER, B. Participant observation and interviewing: A comparison. *Human Organization*, 1957, 16, 28–32.

DUNNETTE, M. D. Personnel management. *Annual Review of Psychology*, 1962, 13, 285–314.

DUNNETTE, M. D., & HENEMAN, H. G. Influence of scale administrator on employee attitude responses. *Journal of Applied Psychology*, 1956, 40, 73–77.

HECHINGER, F. M. A survey of surveys. *New York Times*, September 11, 1966, p. 7E.

KAHN, R. L., & CANNELL, C. F. *The dynamics of interviewing*. New York: Wiley, 1957.

KLEIN, S. M., MAHER, J. R., & DUNNINGTON, R. A. Differences between identified and anonymous subjects in responding to an industrial opinion survey. *Journal of Applied Psychology*, 1967, 51, 152–160.

MASLOW, A. H. *The psychology of science*. New York: Harper & Row, 1966.

PELZ, D. C. The influence of anonymity on expressed attitudes. *Human Organization*, 1959, 18, 88–91.

RICHARDSON, S. A., DOHRENWEND, B. S., & KLEIN, D. *Interviewing: Its forms and functions*. New York: Basic Books, 1965.

WHYTE, W. F. Interviewing in field research. In R. N. Adams & J. J. Preiss (Eds.), *Human organization research*. Homewood, Ill.: Dorsey Press, 1960. Pp. 352–374.

(Received September 20, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 4, 341–342

ADDENDUM TO "AN EMPIRICAL INVESTIGATION OF TWO IMPLICATIONS OF THE TWO-FACTOR THEORY OF JOB SATISFACTION"

GEORGE B. GRAEN AND CHARLES L. HULIN

University of Illinois, Urbana

Multiple hypotheses were tested between the 2-factor theory of Herzberg, Mausner, and Snyderman and the traditional theory of job satisfaction. A procedure was used that granted the assumption of the 2-factor theory of a dimensional difference between overall satisfaction and dissatisfaction. Instruments employed were designed to capitalize on any unique variance associated with the satisfaction or the dissatisfaction dimension. Analysis consisted of 1-way analyses of variance of the Job Description Index variables of Work Itself, Promotion, Salary, Supervision, and Co-workers on overall satisfaction and on overall dissatisfaction. Results support the hypotheses of the traditional theory at the expense of the 2-factor theory.

Attempts to test the two-factor theory of job satisfaction (Herzberg, 1966; Herzberg, Mausner, & Snyderman, 1959) employing methods other than storytelling, in addition to failing to support hypotheses derived from the theory, have assumed a single continuum of overall job satisfaction-dissatisfaction. The single exception was a recent study by Hulin and Smith (1967) in which special instruments were designed to measure independently satisfaction and dissatisfaction. The purpose of the present study was to examine these unique data from the perspective of *strong inference* (Platt, 1964). The strong-inference approach pits hypotheses from competing theories against one another and demands a determination of their relative value for future research. It differs from the more usual approach by emphasizing hypotheses testing rather than theory confirmation (Dunnette, 1966).

Hypotheses derived from the two-factor and the traditional theory (Ewen, Smith, Hulin, & Locke, 1966; Graen, 1966) and tested in this study are as follows. The two-factor theory hypothesizes that those variables classified as "satisfiers" (Work Itself and Promotion) should control a significant amount of variance in overall satisfaction and *not* in overall dissatisfaction, whereas those variables classified as "dissatisfiers" (Co-workers, Supervision, and Salary) should control a significant amount of variance in overall dissatisfaction and *not* in overall satisfaction. In contrast, traditional theory hypothesizes that both so-called "satisfiers" and "dissatisfiers" should control a significant amount of variance in overall satisfaction *and* overall dissatisfaction.

METHOD

The *Ss* were male and female home-office employees of one corporation. Each *S* completed the Job Description Index (Smith, 1967) and either an overall job satisfaction or overall job dissatisfaction instrument. The Job Description Index (JDI) measures satisfaction with the job variables of Work Itself, Promotion, Co-workers, Supervision, and Salary. In this study, 146 *Ss* completed the overall satisfaction measure, and another 155 completed the overall dissatisfaction measure. For further details, see Hulin and Smith (1967).

The analysis consisted of performing one-way analyses of variance to test the effects of the five JDI variables on overall satisfaction and on overall dissatisfaction (Winer, 1962). Each JDI variable was trichotomized into dissatisfied, neutral, and satisfied using cutting points supplied by Ewen (1965). Omega-squared (Hays, 1963) was employed to estimate the strength of relationships between JDI variables and the two criteria of overall satisfaction and overall dissatisfaction.

RESULTS

Those variables classified as satisfiers (Work Itself and Promotion) showed significant relationships on satisfaction *and* on dissatisfaction. In contrast, those variables classified as dissatisfiers (Co-workers, Supervision, and Salary) showed significant relationships on satisfaction and *not* on dissatisfaction. The percentage of variance in satisfaction accounted for by the job variables was 23%

TABLE 1

ANALYSIS OF VARIANCE AND OMEGA-SQUARE OF JOB VARIABLES ON SATISFACTION AND DISSATISFACTION

Independent variable	Dependent variable					
	Satisfaction only			Dissatisfaction only		
	df	*F*	%*	*df*	*F*	%
Satisfier						
Work itself	2	22.28*	.23	2	5.35*	.06
Promotion	2	14.05*	.16	2	8.17*	.09
Dissatisfier						
Co-workers	2	11.85*	.14	2	1.07	.01
Supervision	2	9.81*	.11	2	1.20	.01
Salary	2	5.46*	.06	2	1.32	.01
Residual	143			152		

* % indicates the proportion of total variance accounted for by the independent variable.
* $p \leq .01$.

for Work Itself, 16% for Promotion, 14% for Co-workers, 11% for Supervision, and 6% for Salary. The percentage of variance in dissatisfaction accounted for by the job variables was 6% for Work Itself, 9% for Promotion, and a maximum of 1% each for Co-workers, Supervision, and Salary.

DISCUSSION

Even when the assumption of a dimensional difference between satisfaction and dissatisfaction is made and instruments designed to capitalize on any unique variance associated with each dimension are employed, the results clearly *disconfirm* predictions of the two-factor theory. The findings that "satisfier" variables (Work Itself and Promotion) contribute to both satisfaction and dissatisfaction and that "dissatisfier" variables (Co-workers, Supervision, and Salary) contribute to satisfaction and not dissatisfaction are incompatible with the two-factor theory.

The present results are more compatible with the traditional theory. The traditional theory hypothesizes that if job variables are categorized according to the two-factor classification system, both so-called satisfiers and dissatisfiers contribute over the entire continuum of overall job satisfaction-dissatisfaction. In addition, the traditional theory accounts for the differences on overall satisfaction-dissatisfaction between satisfiers and dissatisfiers by hypothesizing that intrinsic variables (satisfiers) are related more strongly to overall satisfaction-dissatisfaction than are extrinsic variables (dissatisfiers). The traditional theory views the differences between satisfiers and dissatisfiers as primarily a matter of potency rather than a matter of directionality. The only finding not supporting the traditional theory was the failure of the dissatisfier variables to contribute to dissatisfaction. This lack of relationship was probably due to the severe restriction of range in the measure of overall dissatisfaction. In future studies with samples representing lower levels of dissatisfaction, higher relationships would be expected between all job variables and overall dissatisfaction. The results of this study must be viewed as damaging to the two-factor theory and as supporting the traditional theory.

REFERENCES

DUNNETTE, M. D. Fads, fashions, and folderol in psychology. *American Psychologist*, 1966, **21**, 343–352.

EWEN, R. B. Weighting components of job satisfaction. Unpublished doctoral dissertation, University of Illinois, 1965.

EWEN, R. B., SMITH, P. C., HULIN, C. L., & LOCKE, E. A. An empirical test of the Herzberg two-factor theory. *Journal of Applied Psychology*, 1966, **50**, 544–550.

GRAEN, G. B. Addendum to "An empirical test of the Herzberg two-factor theory." *Journal of Applied Psychology*, 1966, **50**, 551–555.

HAYS, W. L. *Statistics for psychologists*. New York: Holt, Rinehart & Winston, 1963.

HERZBERG, F. *Work and the nature of man*. New York: World, 1966.

HERZBERG, F., MAUSNER, B., & SNYDERMAN, B. B. *The motivation to work*. (2nd ed.) New York: Wiley, 1959.

HULIN, C. L., & SMITH, P. A. An empirical investigation of two implications of the two-factor theory of job satisfaction. *Journal of Applied Psychology*, 1967, **51**, 396–402.

PLATT, J. R. Strong inference. *Science*, 1964, **146**, 347–352.

SMITH, P. C. The development of a method of measuring job satisfaction: The Cornell studies. In E. A. Fleishman (Ed.), *Studies in personnel and industrial psychology*. Homewood, Ill.: Dorsey Press, 1967.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received September 28, 1967)

journal of applied psychology

VOL. 52, NO. 5 OCTOBER 1968

A PSYCHOPHYSICAL DETERMINATION OF EQUITABLE PAYMENT:

A METHODOLOGICAL STUDY [1]

SHELDON ZEDECK AND PATRICIA CAIN SMITH

Bowling Green State University

The practicality of using psychophysical methods to determine ranges for payment plans was explored. Thresholds of perceived equitable payment (PSE) and the "just meaningful difference" (jmd) of payment were determined by an adaptation of the psychophysical Method of Limits. PSE and jmd were significantly greater for junior executives than for secretaries of an academic institution, whereas the respective Weber ratios, *K* (the proportionate meaningful additions to the base salary), were not significantly different. There were no significant differences between 2 secretarial subgroups in jmd and *K*, but a significant difference in PSE. The limitations and implications of this method, and its relevance to equity theory, were discussed.

Several theorists have independently proposed models for the determination of equity of payment for work (Adams, 1963, 1965; Homans, 1961; Jaques, 1961; Patchen, 1961; Sayles, 1958; Zaleznik, Christenson, & Roethlisberger, 1958). Each model assumes that all employees perceive a fair, just, or equitable return for what they contribute to their jobs. Furthermore, each model involves a social comparison (an employee determines what his equitable return should be after comparing his inputs and outcomes, or his investments and rewards, with those of co-workers or reference groups). Also, each model assumes, implicitly, that the dissonant cognition by the employee of over- or undercompensation results in feelings of tension, dissatisfaction, anger, or guilt.

Methods of assessing and evaluating equitable payment differ. Typical methods include interviews and discussions with employers and employees, observations by investigators, and the collection of biographical data. Adams' (1963, 1965) extensive research characteristically induced the feelings of under- and overcompensation and then evaluated *Ss*' production records.

A premise that has not been directly investigated is that an employee can evaluate his income as equitable or inequitable in terms of some composite internal absolute standard which is not tied to any one group or person. A further related problem is the sensitivity of the worker to salary inequities. Both Jaques (1961) and Adams (1965) suggest that workers are more sensitive to undercompensation than to overcompensation. Andrews (1967) attempted to reflect this idea in the hypothesis that a small underpayment will affect job performance as much as or more than a large overpayment. The hypothesis was partially confirmed, yet there still is no direct, reliable evidence for the differential sensitivity hypothesis.

The present study investigated the intuitive internal perceptions of salary. Also, the study investigates directly the sensitivity of workers to inequities of pay. To determine the upper and lower thresholds of perceived equitable payment and the "just meaningful difference" (jmd) of payment (analogous to the jnd in psychophysical measurement), the study used

[1] This research is based in part on a master's thesis at Bowling Green State University.

© 1968 by the American Psychological Association, Inc.

an adaptation of the Method of Limits which is the one method that most clearly defines the concept of thresholds (Woodworth & Schlosberg, 1954, p. 199). This method also makes it possible to determine amounts of proportionate meaningful increases (K) to base salary.

METHOD

Subjects

Questionnaires were mailed to 25 male junior executives (Group I), and 96 female secretaries—45 in the maintenance department (Group II) and 51 in the executive department (Group III)—of a large midwestern academic institution. Each had at least 1 yr. of experience in his position.

Returns of the questionnaire were 80% ($n = 20$) for Group I, 51% ($n = 23$) for Group II, 45% ($n = 23$) for Group III, and 6 from the secretarial pool for whom the secretarial office could not be identified.

Research Material

The questionnaire contained general job-satisfaction questions, intended to disguise the intent of the study, and the basic questions necessary for the psychophysical analysis. The format of the basic question was:

Considering the type of work *required* of you on your present job, is _____ a year

_____ a fair salary
_____ more than a fair salary
_____ less than a fair salary.

The salary levels within the basic question (one level, randomly placed, per page) were based on the average salary and probable error (PE) for the respective groups. There were nine salary levels, four levels above and four levels below the average. Each successive change above and below the average was an approximation of the PE: $180 and $720 a year for secretaries and executives, respectively.

Procedure and Design

The point of subjective equity (PSE), or the salary that the S perceived to be fair, and jmd were determined for each group by the Method of Limits with the adaptation of presentation of single stimuli in haphazard order (Guilford, 1954; Woodworth & Schlosberg, 1954).

Two criteria were used in evaluating the data matrices (on which the responses were recorded in serial order). First, if the responses within an employee's questionnaire did not contain an "=" (fair), "+" (more than fair), and "–" (less than fair), then that questionnaire was not used in the analysis. Second, one inversion within an employee's results was tolerated. If more than one inversion occurred, then the following rules were applied: If in inspecting the series in ascending order, a – response was found after there had been a series of – and at least two = responses, or an = and a +, such a series was not evaluated. Similarly, in descending inspection, the appearance of a + after two =s, or an = and a –, disqualified the questionnaire.

TABLE 1

PSYCHOPHYSICAL DETERMINATIONS

Group	Psychophysical measure	n	df	t
	PSE			
I	$7920.00	15		
II & III	$3814.29	21	34	18.42*
II	$3574.29	7		
III	$3919.09	11	16	2.87*
	jmd			
I	$ 912.00	15		
II & III	$ 235.71	21	34	6.19*
II	$ 295.71	7		
III	$ 196.36	11	16	1.30
	K			
I	.115	15		
II & III	.062	21	34	.58
II	.083	7		
III	.050	11	16	.29

* $p < .05$.

RESULTS

After applying the criteria for evaluation, the sample sizes for Groups I, II, and III were reduced to 15, 7, and 11, respectively, with 21 secretaries forming the combined secretarial group (including 3 unclassified secretaries).

In comparison of the junior executives with the total secretarial pool and with each sample of secretaries, both the PSE and the jmd were significantly greater for the executives (Table 1). In contrast, however, a comparison of the respective Weber ratios reveals that the proportionate meaningful additions to the base salary were neither large nor significantly different.

A comparison of the two secretarial groups showed that although the actual earned salaries for Groups II and III were within the same range, the PSE was significantly greater

TABLE 2

PERCEIVED MEAN SALARY RANGES

Group	n	<Fair	Equitable (PSE ± jmd)	>Fair
Junior executives	15	$7008.00	$7008.00–8832.00	$8832.00
Secretaries	21	$3578.58	$3578.58–4050.00	$4050.00
Group II	7	$3278.58	$3278.58–3870.00	$3870.00
Group III	11	$3722.73	$3722.73–4115.45	$4115.45

for Group III, whereas the jmd was not significantly different. The difference between K was also small and not significantly different.

Based on this information, equitable pay ranges for all groups were constructed (Table 2).

Since the sample size for each group was small, the combined sample of returned questionnaires was analyzed (Table 3). The lower threshold, T(−), for each employee who had indicated an upper threshold, T(+), was compared with the T(−) of those employees who had not stated a T(+). For both executives and the total secretarial group, the T(−) was significantly smaller for the subgroups that had indicated a T(+).

TABLE 3

LOWER THRESHOLD, T(−), FOR GROUPS WITH AND WITHOUT UPPER THRESHOLD, T(+)

Group	With T(+)	Without T(+)	df	t
Junior executives	$7008.00 ($n$ = 15)	$9150.00 ($n$ = 5)	18	3.67*
Secretaries	$3578.57 ($n$ = 21)	$3893.48 ($n$ = 23)	22	3.01*

* $p < .05$.

DISCUSSION

This study demonstrates the potential utility of this method, despite the small sample size. The results of this study are, of course, situation specific; the PSE and jmd that were determined apply only to the groups on which the data were collected. The results cannot be generalized to a similar work group in the same setting (as evidenced by the differences obtained with the two secretarial groups), much less to a similar group in a different setting. The method may, however, have general applicability.

It is to be expected that, as an employee goes up the work-level scale, PSE and jmd will increase. However, the nature of the functional relationships involved is an interesting point for speculation based on the data obtained. Since there was no significant difference between the K values for the groups, either Weber's, Fechner's (Guilford, 1954), or Stevens' (1957) law might hold. Applying this technique to a greater number of groups along the work and pay-grade scale would reveal the nature of the wage curves in operation for any specific company.

The technique also detects differences between groups at the same job level. The significant difference between PSE for the two secretarial groups could be attributed to different working conditions for the two groups, to differences in personal needs, or other causes which can be subsequently investigated.

Though the two secretarial groups had different perceptions of fair salaries, the amount of meaningful increase was approximately the same. It is not suggested that the average K of 6.2% should be applied in determining raises, but that a range such as 5–8% be used. This range would be useful in determining raises with consideration for education, experience, seniority, etc. The use of ranges should be applied for all determinations based on the method, instead of using the exact PSE, jmd, or K, and could correct for the individual differences within the groups.

A point to be considered in using this technique is that of sensitivity to inequities. Adams (1965), Jaques (1961), and Andrews (1967) suggest that Ss are more sensitive to underpayment as opposed to overpayment. The data are not inconsistent with these generalizations. Since the choices included four PEs above and four PEs below the average salary, one concludes that Ss were more willing to indicate inequities below than above equitable salaries. For the junior executives, 100% of the sample indicated a T(−) whereas only 75% indicated a T(+). The results were more striking with the secretaries, 98% of whom indicated a T(−) while only 49% indicated a T(+). However, these findings cannot be taken as conclusive for it is possibly an artifact of the number and values

of the levels used, and might also indicate dissatisfaction.

Since more than 50% of the returned questionnaires did not meet the criteria for evaluation, an attempt was made to extract some information from the unevaluated questionnaires by comparing the T(−) responses of the unevaluated with the evaluated questionnaires. The significant difference between the T(−) responses for both the executives and secretaries could be interpreted as indicating that the differences were only in intervals of uncertainty (equitable range) and not a difference in jmd.

Limitations

Several technical limitations must be considered. The results were obtained by a mailed questionnaire. This method of data collection is hazardous for this type of study. It is important, for validity purposes, that *S* fills out one page at a time and does not look back to his previous answers. The *S* should not have the opportunity to scan the questionnaire, familiarize himself with the levels, and then, with the aim of being consistent, fill in his responses. A suggested procedure is to administer the questionnaire to a group of employees simultaneously, requesting each *S* to answer one page at a time, hand in or place a completed page in an envelope, and then receive another page, the process being repeated until the task is completed. Also, with this type of administration, the sample size would be increased.

Another deficiency in the technique is the number and size of increments. The value of the increments should be sufficiently small so as to be most sensitive to the inequities. The desired sensitivity can be gained by following a test-retest procedure. After administering the questionnaire with nine levels of salary, one could retest the same group a short time after with the same filler questions, but with nine different, though appropriate, salary levels, resulting in 18 levels from which to determine PSE, jmd, and *K*.

Implications

Once the technique is refined, general information as to who perceives himself fairly paid, less than fairly paid, and more than fairly paid will be provided. On the basis of these determinations, several of the hypotheses that Adams and his associates have tested could be examined without inducing feelings of inequity. The quantity and quality of work for each employee could be compared and differences, if any, between those perceiving themselves equitably and inequitably paid evaluated. The influence of the input variables on perceptions could be examined; for example, what is the jmd for someone with 20 yr. of experience as opposed to someone with 5 yr. of experience? The effects of task differences on perceived equity for groups with similar salary scales could be investigated.

Another analysis could be made of the relationship of a person's perceptions to his satisfaction. It is conceivable that anyone who would deviate from the equitable interval would also be dissatisfied with his job and its conditions.

Still another interesting analysis would be the determination of jmd for males as opposed to females in the same work group, or career women as opposed to women supplementing their families' incomes.

Finally, it would be possible to use the results when considering the granting of raises and increases due to promotions, or for distinguishing different levels of jobs. It appears that this would be an improvement over the arbitrary techniques that are presently being used in industry. As Jaques (1961) pointed out, equitable payment should prove economically sound as well as socially just, because it would diminish wrangling and encourage efficiency. Furthermore, Opsahl and Dunnette (1966) found in their review of the role of financial compensation in industrial motivation that companies "assign" each employee a "pay curve" which is the result of successive alterations in compensation and compensation policies through the years. They noted that the usual way of doing this is with little or no advanced planning. Increments are simply given haphazardly on a year-to-year basis and the resulting career curve is equally haphazard. If this is the situation, if increments are given without planning and regard for its meaning to the employee, then a better technique is needed.

REFERENCES

ADAMS, J. S. Toward an understanding of inequity. *Journal of Abnormal and Social Psychology*, 1963, **67**, 422–436.

ADAMS, J. S. Injustice in social change. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 2. New York: Academic Press, 1965.

ANDREWS, I. R. Wage inequity and job performance: An experimental study. *Journal of Applied Psychology*, 1967, **51**, 39–45.

GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.

HOMANS, G. C. *Social behavior: Its elementary forms*. New York: Harcourt, Brace & World, 1961.

JAQUES, E. *Equitable payment*. New York: Wiley, 1961.

OPSAHL, R. L., & DUNNETTE, M. D. Role of financial compensation in industrial motivation. *Psychological Bulletin*, 1966, **66**, 94–118.

PATCHEN, M. *The choice of wage comparisons*. Englewood Cliffs, N. J.: Prentice-Hall, 1961.

SAYLES, L. R. *Behavior of industrial work groups: Prediction and control*. New York: Wiley, 1958.

STEVENS, S. S. On the psychophysical law. *Psychological Review*, 1957, **64**, 153–181.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, Rinehart & Winston, 1954.

ZALEZNIK, A., CHRISTENSON, C. R., & ROETHLISBERGER, F. J. *The motivation, productivity, and satisfaction of workers: A prediction study*. Boston: Harvard University, Graduate School of Business Administration, 1958.

(Received October 24, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 5, 348–356

SEARCHING FOR LETTERS OR CLOSED SHAPES IN SIMULATED ELECTRONIC DISPLAYS[1]

E. C. POULTON

Applied Psychology Research Unit, Cambridge, England

16 men counted the number of bright symbols of a particular kind in a simulated electronic display presenting a total of between 19 and 59 symbols of 8 different kinds. In some displays all the symbols were bright, in others bright and dim symbols were mixed. 4 sets of 8 symbols were compared. Targets were missed less often from sets of capital letters than from sets of closed shapes ($p < .01$); thus the letters were the more easily discriminated from each other. The probability of missing at least 1 target increased approximately linearly with the number of targets present ($p < .01$). The rate of increase was approximately proportional to the total number of bright symbols displayed, and was somewhat greater when dim symbols were present also ($p < .05$). For a combined total of 57–59 bright and dim symbols, the probability of missing at least 1 target increased from about .03 per target when 10 of the symbols were bright, to .09 when 26 were bright. 35% of the underestimations were of size 2 or greater. The time taken to count the targets increased with the total number of bright symbols present ($p < .01$). Dim symbols tended to be counted as bright ($p < .01$).

Electronic display systems are capable of showing the locations of large numbers of objects at a time on a single cathode-ray tube. Where the objects can be classified into classes, the objects belonging to a class can all be represented by the same symbol. For example, a different capital letter can be used for each class. Since part of the man's difficulty in spotting items of a particular class consists in discriminating between them and the remaining items, the letters used should be easily discriminable from each other (Howell & Kraft, 1959). Men working with displays of this kind may be accustomed to using simple shapes such as squares, triangles, and diamonds; in which case they are not likely to relish a change to capital letters unless letters can be shown to be more easily discriminable. The principal aim of the experiment was to compare the discriminability of a set of capital letters with a set of closed shapes. Symbols intermediate between letters and shapes were also included (see Figure 1) in an attempt to determine the characteristics of letters and shapes which help or hinder discriminability. No experimental work comparing the relative discriminability of letters and shapes appears to have been published previously.

In an experiment of this nature additional independent variables are the number of symbols of the class for which the man is searching, and the total number of symbols displayed (Sleight, 1952). If the symbols are presented bright against a background of additional dim symbols of similar shapes, the number of dim symbols can be another independent variable. A subsidiary aim of the experiment was to determine the effect of these three distinct sources of display clutter upon the man's ability to spot target symbols.

METHOD

Materials

The four sets of symbols are illustrated in Figure 1. The symbols were about 2.7 mm. tall, which is the height of 12-point newspaper capitals. The shapes in Column 4 were chosen to be representative of the kinds of shapes to which controllers might have become accustomed during the course of their duties. The capital letters in Column 1 were selected to give a set of shapes with features similar to those of the shapes in Column 4 when the spaces within the letters were enclosed as in Column 3. Only one shape, the triangle with the apex at the top, is common to Columns 3 and 4. But both columns contain three pairs of shapes which can be confused (see Table 2). Column 2 represents a degree of letter

[1] The research was carried out in 1960 with the financial support of the British Medical Research Council. D. C. V. Simmonds and S. G. Blake-Smith very kindly made the slides. Statistical advice from P. M. E. Altham is also gratefully acknowledged.

closure intermediate between those in Columns 1 and 3.

Twelve basic spatial patterns of symbols were selected, 2 for practice and 10 for the experiment proper. Both the practice patterns and 6 out of the 10 test patterns contained both bright and dim symbols. The remaining four test patterns contained only bright symbols. Each of the 12 basic patterns was rotated once through 90°, once through 180°, and once through 270°, to give altogether 48 arrangements of symbols, 8 for practice and 40 for the experiment. Thus all the practice arrangements and 24 of the test arrangements contained both bright and dim symbols, while the remaining 16 test arrangements contained only bright symbols.

The 48 arrangements were reproduced four times, once in each of the four sets of symbols shown in the columns of Figure 1. The symbols were white on a black background as in the figure. To produce both bright and dim symbols on the same photographic negative, the display was photographed twice by a clamped camera without changing the plate. First a short exposure took all the symbols. Then the symbols which were to be dim were removed, and a second exposure was made of the symbols which were to be bright. The pictures were finally reduced in size so that the positives could be mounted in 2 × 2 in. slide cases.

For each of the 48 arrangements the symbols in one of the rows of Figure 1 were selected as the targets. In the 8 practice arrangements of a set, each of the 8 kinds of symbol in the set was selected once. The number of bright target symbols on the slides ranged from 0 to 6. In the 24 test arrangements with both bright and dim symbols, each of the 8 kinds of symbol in the set was selected as the target 3 times. In the 16 test arrangements with only bright symbols, each of 8 kinds of symbol was selected twice. Selection was quasi-random with the restriction that the same symbol was not a target in two tests derived from the same basic pattern. The number of bright target symbols on the slides ranged from 0 to 11. Where both bright and dim symbols were present, the total number of bright symbols ranged from 10 to 26, the total number of dim from 11 to 47. Where only bright symbols were present, their numbers ranged from 19 to 55.

Experimental Design and Subjects

For each of the four experimental conditions corresponding to the four sets of symbols in Figure 1, the 40 experimental slides were arranged in four lots of 10, each lot containing 1 of each of the 10 basic patterns. The order of the four lots, and the order of the arrangements within lots, was the same for each experimental condition.

Sixteen men acted as experimental *S*s. Their ages ranged from the mid 20s to the late 50s. One group of four men was experienced with electronic displays of this nature, another group of four men was very experienced, while the other two groups were, respectively, not very experienced and inexperienced. Each man in a group of four received the four experimental conditions corresponding to the sets of symbols in Figure 1 in a different order, according to one of four separate Latin-square designs.

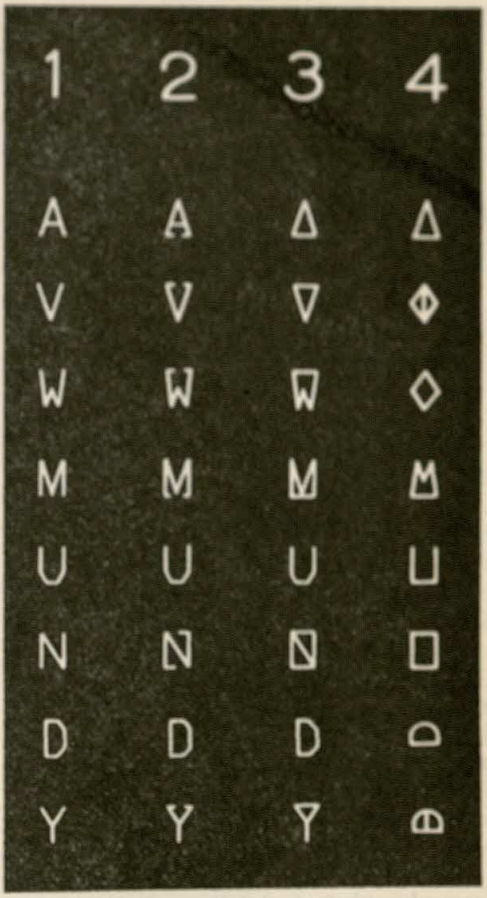


FIG. 1. The four sets of symbols compared. (Column 1: letters; Column 2: semiclosed letters; Column 3: closed letters; Column 4: shapes.)

Procedure

Each man was tested separately. He sat in front of a vertical semitransparent circular screen of 12 in. diameter, upon the back of which the slides were projected. The center of the screen was approximately level with his eyes, and about 18 in. away. He was told that he could move the screen if desired to focus the image of a slide, and to center it.

Before each slide was presented the man was told which was to be the target symbol, and whether or not the slide would contain dim symbols as well as the bright ones which alone had to be looked at. The slide was not presented until he indicated that he was ready to begin. As soon as the slide appeared, he had to count all the bright target symbols and call out his answer. He was told to make sure

that his answer was correct before he gave it, but not to waste time unnecessarily. A stop watch was started as each slide appeared, and stopped as soon as he gave his answer. While the answer and time taken were being recorded by the *E*'s assistant, the man was told which was to be the next target symbol.

After each of the eight practice slides the man was told whether his answer was right or not. If he was wrong, *E* went over the slide with him until he discovered his mistake. No knowledge of results was given after any of the 40 test slides. At the end of the session the man was thanked and generally encouraged. He was asked which symbols he confused with which and his preferences for individual symbols and sets of symbols. Including instruction and practice, the first session took an average of just under 30 min., subsequent sessions just under 20 min. Normally each man performed one condition per day, or one condition in the morning and a second in the afternoon. But two men worked alternately one afternoon, since this was the only time at which they were available.

Scoring and Calculations

Errors were classified either as underestimations of the correct number of target symbols or as overestimations. Since over half the errors were only one different from the correct answer, the sizes of the errors have been neglected in most of the quantitative analyses. Two-tailed statistical tests have always been used except where stated.

RESULTS

Differences between Sets of Symbols

Incidence of underestimations. Table 1 gives the incidence of errors separately for the four sets of symbols, both when only bright symbols were present in the display, and when dim symbols were present as well as bright. Answers underestimating the correct number of target symbols are shown separately from overestimations. The data in the bottom row of the table, pooled over all four sets of symbols, show that the incidence of underestimations was virtually unaffected by the presence of dim symbols in the display. In comparing the four sets of symbols on underestimations, the results from all 40 slides have therefore been pooled, as shown in the right-hand column of the table.

Analysis of variance on the pooled underestimations indicated that the effects of sets of symbols and of practice were both highly reliable ($p < .001$). Differences between individuals were also highly reliable, but the small differences between the groups of men with different degrees of experience of electronic displays were not reliable. On Tukey's method for multiple comparisons (Ryan, 1959), both the letters and the semiclosed letters gave reliably fewer underestimations than either the closed letters or the shapes ($p < .05$ or better).

As a result of the Latin-square design, the shapes could have been at a disadvantage if there were greater positive transfer between the three sets of symbols resembling letters, than from these three sets to the shapes. To check that the underestimations of the shapes in the last column of Table 1 were not due entirely to this cause, the performances were

TABLE 1

MEAN PERCENTAGE OF INCIDENCE OF UNDERESTIMATIONS AND OVERESTIMATIONS ACCORDING TO SYMBOLS DISPLAYED

Set of symbols	16 slides with only bright symbols		24 slides with bright and dim symbols		All 40 slides
	Under-estimations	Over-estimations	Under-estimations	Over-estimations	Under-estimations
Letters	16.4	.4[a]	12.0	12.3[b]	13.8[c]
Semiclosed letters	11.3	3.5[a]	16.4	1.6	14.4[d]
Closed letters	19.2	1.2	23.2	3.1	21.6
Shapes	28.9	.8	21.4	3.6	24.4
All 4 sets of symbols combined	19.0	1.5	18.2	5.1	18.5

[a] Letters-semiclosed letters $p < .02$.
[b] Letters reliably different from all other sets of symbols at the .01 level or better.
[c] Letters-closed letters or shapes $p < .01$.
[d] Semiclosed letters-closed letters or shapes $p < .05$ or better.

examined of the six men whose first two tests involved shapes and either letters or semiclosed letters. Two men first searched for shapes and then for semiclosed letters, while two other men had these two conditions in the reverse order. There were also a pair of men whose orders of searching for shapes and letters balanced each other. Of these six men, five gave more underestimations on the shapes. The sixth gave the same number of underestimations on the semiclosed letters as on the shapes, but had the shapes as the second condition. Correcting for order, the difference between the shapes and the letters or semiclosed letters is reliable on a two-tailed sign test ($p < .05$). Thus the greater number of underestimations of the shapes in Table 1 is unlikely to have been due entirely to cumulative transfer effects.

Incidence of overestimations and reported confusions. The bottom row of Table 1 shows that overestimations were over three times as common on the slides carrying both bright and dim symbols, as on the slides carrying only bright symbols. A Wilcoxon test (Siegel, 1956, pp. 75–83) indicated that the difference was reliable ($p < .01$). The two sets of slides thus need to be treated separately on this measure.

The slides with only bright symbols produced a total of only 15 overestimations. A sign test indicated that there were reliably fewer overestimations with the letters than with the semiclosed letters ($p < .02$). None of the other differences between sets of symbols was reliable here. All but 1 of the 15 overestimations were just one greater than the correct number. They could have been due to counting a single target symbol more than once, or to counting one wrong kind of symbol as a target. The single remaining overestimation was eight too large; it was probably due to counting W's instead of U's.

On the slides containing both bright and dim symbols, half the overestimations were two or more greater than the correct answer. Wilcoxon tests indicated that there were reliably ($p < .01$ or better) more overestimations on the series of slides with letter symbols than on any of the other three series, which did not differ reliably from each other ($p > .05$). Since the letters gave if anything the fewest overestimations when only bright symbols were displayed, most of these overestimations with the combined bright and dim symbols must have been due to counting dim symbols as bright.

TABLE 2

COMMONLY REPORTED CONFUSIONS BETWEEN SYMBOLS

Set of symbols	Pairs confused
Letters	W and M
Semiclosed letters	V and Y
Closed letters	Closed W and M Closed V and Y Closed V and A
Shapes	Triangle and diamond Diamond and split diamond D on side and split D

According to the reports of the experimental *S*s, the most common confusions between symbols were those indicated in Table 2.

Time taken. The mean time taken to search for all the bright target symbols and to call out the number counted was 11.9 sec. on the slides with only bright symbols, compared with 11.4 sec. on the slides with both bright and dim symbols. This difference was reliable on a Wilcoxon test against the differences between experimental *S*s ($p < .02$), but the difference was not reliable when tested against the differences between slides. (As will be pointed out later, the mean response time depended upon the number of bright symbols displayed. The group of slides with only bright symbols happened to include 4 slides with very large numbers of bright symbols, as well as 12 slides with average numbers—see Figure 3.) Separate analyses of variance on the two sets of slides indicated that the only other reliable effect was produced by the differences between individuals ($p < .001$). Mean time taken was not affected reliably by the set of symbols used, by practice during the experiment, nor by previous experience with electronic displays.

Sources of Clutter in the Display

Incidence of underestimations. Figure 2 shows how the incidence of underestimations

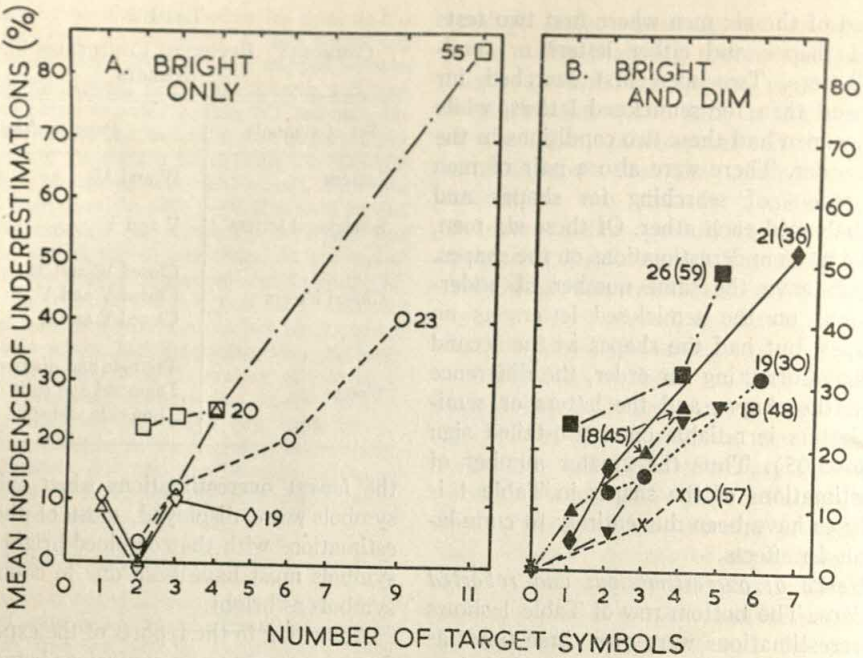


FIG. 2. Mean percentage of incidence of answers underestimating the number of target symbols, shown according to the number of target symbols actually present. (In A for slides with only bright symbols, the numbers printed beside the functions indicate the total number of symbols on the slide. In B for slides with both bright and dim symbols, the first number indicates the number of bright symbols on the slide, the number in brackets indicates the total number of combined bright and dim symbols present. Most points represent data from the same 16 men pooled over 4 slides, one with each set of symbols. But in functions with only 3 points, one of the points represents data pooled over 8 slides.)

of the correct number of target symbols was influenced both by the number of target symbols to be counted, and by the total number of bright symbols displayed. Each function represents one basic pattern of symbols (see Method section). The number printed against the function gives the total number of bright symbols in the basic pattern. In Figure 2B, for the slides in which dim symbols were displayed as well as bright, the number in brackets gives the total number of symbols in the basic pattern, both bright and dim. The four points on the function are for the four different orientations of the basic pattern. Usually the number of signals of the kind designated as target was different for each orientation, but five of the functions have only three points because the number of target signals was the same for two orientations. The data for the four sets of symbols have been pooled. The orientation of the basic pattern of symbols, the particular symbol designated as the target, and the serial position of the slide in the set of 40 test slides have been treated as random variables in this analysis. They may account for some of the residual variability displayed by the data.

Each function in Figure 2 was fitted by a straight line using the method of least squares. In this first analysis the three points at the origin in Figure 2B were excluded, since it is not possible to underestimate the number of targets when none is displayed, yet any function with a point at the origin is almost bound to show a positive slope. The computed slopes of all 10 functions were found to be positive ($p < .01$). This means that as more target symbols were added to the display, the probability of failing to report at least one target symbol increased.

Each of the 10 functions was therefore

fitted by a straight line passing through the origin, again using the method of least squares. The slopes found on this second analysis are given in Table 3. The fitted slopes of the functions in Figure 2A ranged from 2.1% underestimations per target symbol for a total of 19 symbols to 7.6% per target symbol for a total of 55 symbols. The tau coefficient of rank correlation (Siegel, 1956, pp. 213–223) between the fitted slopes and the total number of symbols displayed was +.67. Thus the probability of missing at least one target symbol tended to increase more rapidly with the number of target symbols as the total number of symbols in the display was increased.

The fitted slopes of the functions in Figure 2B were rather steeper, ranging from 2.9 for a total of 10 bright symbols to 9.0 for a total of 26. The four functions for totals of between 18 and 21 bright symbols all had steeper fitted slopes than any of the three corresponding functions in Figure 2A for totals of between 19 and 23 symbols. A Mann-Whitney U test (Siegel, 1956, pp. 116–127) indicated that the average difference was just reliable ($p = .028$ on a one-tailed test). Thus the probability of missing a target symbol increased when dim symbols were added to a display containing only bright symbols.

For the functions in Figure 2B, tau between the fitted slopes and the total numbers of bright symbols in the display was +.60. For the total numbers of both bright and dim symbols in the display, and for the total numbers of dim symbols only, taus were only +.07 and −.07, respectively. Thus the probability of missing a target symbol depended a good deal more upon the total number of bright symbols in the display, than upon the total number of bright and dim symbols combined, or of dim symbols only.

Even the larger values of tau were not reliable statistically ($p > .05$), as there were not enough functions in either of the figures. However where two functions have one or more points in the same vertical columns of the figure, it is possible to compare their relative heights above the abscissa at these points with the number of target symbols held constant. In Figure 2A the functions for 55, 23, and 19 total symbols all have points in the columns for both 3 and 2 target symbols. For

TABLE 3

COMPUTED SLOPES OF FUNCTIONS IN FIGURE 2

Total number of symbols on slide		(Figure 2A) Bright symbols only	(Figure 2B) Bright and dim symbols
Bright	(Bright + dim)		
10	(57)		2.86
18	(45)		6.60[a]
18	(48)		5.43[a]
19		2.13[a]	
19	(30)		5.37[a]
20		5.32[a]	
21	(36)		7.45[a]
23		4.09[a]	
26	(59)		8.99
55		7.61	

Note.—Mean percentage of incidence of underestimations per target symbol.

[a] Of the functions for numbers of bright symbols ranging from 18 through 23, those for dim symbols as well (Figure 2B) all had steeper slopes ($p < .05$ on a one-tailed test).

these two numbers of target symbols combined, the function for the total of 55 displayed symbols lies reliably above the other two functions, which give almost identical results ($p < .02$ or better on Wilcoxon tests). The function for 55 displayed symbols pooled over 4 and 2 target symbols also lies reliably above the function for 20 displayed symbols ($p < .02$).

In Figure 2B toward the bottom, the functions for 10 (57) and 18 (45) total symbols both have points in the columns for both 3 and 4 target symbols. For these two numbers of target symbols combined, the function for 18 (45) displayed symbols lies reliably above the function for 10 (57) displayed symbols ($p < .01$). Toward the top of the figure this same function for 18 (45) displayed symbols pooled over 1 and 4 target symbols lies reliably below the function for 26 (59) displayed symbols ($p < .02$). So does the function for 18 (48) displayed symbols pooled over 4 and 5 target symbols ($p < .01$). Thus the total number of bright symbols in the display clearly affects the heights of the points in Figure 2.

Sizes of underestimations. When two or more target symbols were displayed, underestimations of the correct number of targets

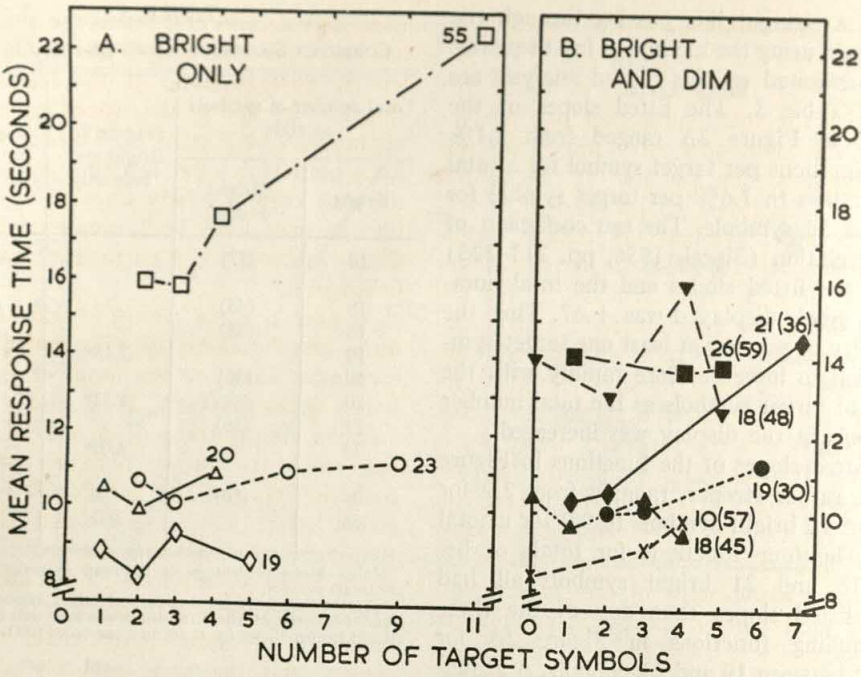


FIG. 3. Mean time taken to count target symbols according to the number of target symbols present. (The parameters are the same as in Figure 2.)

of size 2 and greater averaged 35% of the total number of underestimations. This proportion is a good deal larger than that predicted by a Poisson distribution of underestimations. It strongly suggests that on trials where the man had missed one target symbol, he was likely to miss one or more additional target symbols.

As was to have been expected, the proportion of underestimations of size 2 and greater increased as the number of target symbols increased, from 25% when there were 2 through 4 target symbols to 55% when there were 7 or more target symbols. For the data pooled over *Ss*, the tau correlation between the proportion of underestimations of size 2 and greater and the number (2 or greater) of target symbols displayed was +.56 ($p < .05$).

Time taken. Figure 3 is similar to Figure 2, but the dependent variable on the ordinate is the mean time taken to count all the bright target symbols present, and to call out the number. In Figure 3A only the top function for a total of 55 bright symbols shows an appreciable slope. Here the difference between 11 and 4 target symbols was highly reliable ($p < .001$), and the difference between 4 and both 3 and 2 target symbols was also reliable ($p < .01$). In Figure 3B three of the six functions slope up to the right, while the other three slope down to the right. Clearly, except for the slides with 55 bright symbols, the number of target symbols did not appreciably affect the mean time taken to respond.

However mean response time tended to increase with the number of bright symbols displayed. In Figure 3A the top function for 55 displayed symbols pooled over 3 and 2 target symbols lies reliably above the functions for both 23 and 19 displayed symbols ($p < .001$). Pooled over 4 and 2 target symbols it also lies reliably above the function for 20 displayed symbols ($p < .001$). At the bottom of the figure the function for 19 displayed symbols pooled over 3 and 2 target symbols lies reliably below the function for 23 displayed symbols ($p < .01$). Pooled over 2 and 1 target symbols it also lies reliably below the function for 20 displayed symbols ($p < .001$). The tau correlation between the mean heights of the four functions and the total numbers of

symbols on the slides was +1.00 ($p < .05$ on a one-tailed test).

In Figure 3B the differences in height correlate with the total number of bright symbols on the slide only when there are few bright symbols. The function at the bottom for 10 (57) displayed symbols pooled over 4, 3, and 0 target symbols lies reliably below the function for 18 (45) displayed symbols ($p < .01$). For 3 target symbols it also lies reliably below the function for 19 (30) displayed symbols ($p < .05$). But the relationship between the total number of bright symbols on the slide and the mean time taken does not hold for the functions higher up in the figure. Whereas the function for 18 (45) displayed symbols is one of the lowest, the function for 18 (48) displayed symbols is on average the highest. The tau correlation between the mean heights of the six functions and the total numbers of bright symbols on the slides was +.53. For the total numbers of both bright and dim symbols in the display, and for the total numbers of dim symbols only, tau was only +.07 and −.07, respectively.

The mean response times of the functions in Figure 3A for totals of 19, 20, and 23 bright symbols can be compared with the means of the functions in Figure 3B for totals of 18 (45), 18 (48), 19 (30), and 21 (36) symbols. Both lots of functions are for conditions with about the same total numbers of bright symbols, but the conditions of Figure 3B involved dim symbols also. A Mann-Whitney U test showed no reliable difference between the two lots of functions ($p > .05$). Thus the presence of dim symbols in the display did not appreciably affect the mean time taken to respond.

DISCUSSION

Letters versus Shapes

The right-hand column of Table 1 shows that the number of target symbols displayed was less likely to be underestimated when a set of capital letters was used than when a set of closed letters or shapes was used. A possible reason for the greater efficiency of the set of letters is that the forms of capital letters have been enormously overlearned. Ever since a child first learns the alphabet, he has been perceiving and identifying capital letters, whereas some of the forms of the closed letters and of the shapes are relatively unfamiliar (see Figure 1). Thus in searching for a particular letter among a set of letters, the man knew precisely what he was looking for; he may have been somewhat less sure when searching for a particular closed letter or shape.

Another possible reason is that the letter forms may be more easily discriminable from each other than are the forms of the closed letters or the shapes. V, W, U, and Y all have an open end at the top; A and M have an open end at the bottom; N has an open end at both the top and the bottom; D is a fully closed letter; whereas the closed letters and shapes are all closed with the single exception of the U. It is clear from Howell and Kraft's (1959) confusion matrices that a less confusible set of capital letters could have been chosen, and the same probably applies to sets of closed shapes (Bowen, Andreassi, Truax, & Orlansky, 1960). But in so far as some of the letters are closed and some are open with the openings pointing in different directions, the direction of the opening supplies a cue which is not available with sets of closed forms. This point does not always appear to have been appreciated in considering sets of symbols for electronic displays (Bowen et al., 1960; Sleight, 1952; Vicino, Andrews, & Ringel, 1966).

Cluttered Displays

Figures 2 and 3 illustrate the effect of three types of clutter: total number of symbols displayed, number of bright symbols when only bright symbols require attention, and number of target symbols requiring immediate attention. Figure 2 shows that the probability of missing at least one target symbol increased more or less linearly with the number of target symbols displayed, as if each target symbol had an equal chance of being the first one to be missed, which was independent of the number of other target symbols present. The data for multiple underestimations suggest that once one target symbol had been missed, there was an increased chance of missing at least one more target symbol.

Figure 3 shows that the number of target symbols displayed did not affect the time taken to count them when the total number of bright symbols was 26 or less. But with 55 bright symbols in the display, four or more target symbols did cause extra delay, presumably as a result of recounts.

Figure 2 shows that the probability per target symbol of missing at least one of them increased as the total number of bright symbols in the display increased. Table 3 indicates that for the data of both Figures 2A and B, an increase in the total number of bright symbols by rather less than a factor of 3 is accompanied by an increase by rather more than a factor of 3 in the probability of missing at least one target symbol. The additional bright symbols must either have masked, or else distracted attention from, the target symbols.

Figure 3 shows that it also took longer to count the target symbols when there were more bright symbols in the display. Presumably the man had to examine all the bright symbols to ensure that he did not miss any targets. It follows that the man should be provided with a facility for removing from the display categories of bright symbol which are not relevant to his present purpose.

The chief effect of adding dim symbols to the display was to increase the probability of overestimating the number of bright target symbols (Table 1), some dim symbols being counted as targets. Table 3 shows that adding dim symbols also slightly but reliably increased the chances of missing at least one bright target symbol. Thus if the man does not need to refer to certain symbols, he should be able to remove them completely from the display, rather than being able only to reduce their brightness.

REFERENCES

BOWEN, H. M., ANDREASSI, J. L., TRUAX, S., & ORLANSKY, J. Optimum symbols for radar displays. *Human Factors*, 1960, **2**, 28–33.

HOWELL, W. C., & KRAFT, C. L. Size, blur, and contrast as variables affecting the legibility of alphanumeric symbols on radar-type displays. Wright Air Development Center, Tech. Rep. 59–536, 1959.

RYAN, T. A. Multiple comparisons in psychological research. *Psychological Bulletin*, 1959, **56**, 26–47.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

SLEIGHT, R. B. The relative discriminability of several geometric forms. *Journal of Experimental Psychology*, 1952, **43**, 324–328.

VICINO, F. L., ANDREWS, R. S., & RINGEL, S. Type, extent and coding of updated symbolic information. *Human Factors*, 1966, **8**, 407–416.

(Received September 27, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 5, 357–361

EFFECT OF CONTROL PLACEMENT ON INFORMATION TRANSFER RATE USING BISENSORY SIGNALS

A. D. PERRIMENT

Monash University, Clayton, Victoria, Australia

A 2 × 2 × 4 factorial experiment was designed to examine the effect which various display-control relationships had upon the rate at which *S*s could transmit information from a bisensory display. *S*s were required to respond to a visual signal simultaneously presented with an auditory signal of equal uncertainty. Responses were made by depressing push buttons. The 3 response-code variables examined were (1) the code carried by each of the operating limbs, (2) the code carried by the operating digits of each hand, and (3) the degree of separation between the button pairs. The only variable significantly affecting information transfer rate was found to be the code carried by the operating limbs, although the other 2 variables combined with it to produce significant interactions. The results support the contention that the S-R expectancies, which *S* brings to the task, are important determinants of performance. The flexibility of the human component in man-machine systems is also illustrated.

A previously reported study (Perriment & Webster, 1967) drew attention to some of the factors influencing optimal display-control arrangements, where a bisensory signal required a bimanual motor response. The present paper reports an experiment designed to examine further two of the previous findings, which were held to be of particular interest. One of these was the apparent superiority, in terms of information transfer rate (ITR), of controls compactly arranged and located directly in front of *S*, compared with controls which were laterally separated from the midline. The other was the evidence obtained for complex interaction effects between the three response-code variables studied. The work reported here was directed at examining the generality of the control separation effect, and investigating in greater detail the interaction effects obtained.

METHOD

Apparatus

The apparatus has been described in detail elsewhere (Perriment & Webster, 1967). For the present study, the response panel was modified to permit a graduated lateral displacement of the response buttons. Briefly, the signal display consisted of a 1000-cps tone and a light flash, simultaneously presented through earphones and a pair of lamps located directly in front. The buttons were arranged on the panel in one of four ways: in the form of a rectangle directly in front of *S* (this arrangement is referred to as zero separation), or with each pair laterally separated from the other by 6, 12, or 24 in. The *S*'s task was to record the directional characteristics of the signal components, that is, whether the left or right lamp flashed, and whether the tone occurred at the left or right earphone. As each of the signal components required a binary judgment, the uncertainty associated with each bisensory event was 2 bits. Signals were presented at approximately 5-sec. intervals, and immediately preceded by the onset of a red warning light centrally placed between the two signal lamps. The *E* was screened from *S* and presented the signals in a prearranged order, which was randomized within the constraint of equiprobability of occurrence for each signal type. Responses were recorded manually by *E*. The circuitry of the system allowed the identification of responses to each component of the signal, together with a measure of reaction time (elapsed time between the onset of the signal and the completion of the total response) in milliseconds.

Subjects

Sixty-four volunteer *S*s were drawn from an introductory psychology class. No inducement was offered for participation. The *S*s were allocated to the various treatment groups according to their attendance at the laboratory. None had previous experience of the task required. As the earlier study had failed to produce evidence supporting the hypothesis of sex differences in performance on this task, sex was disregarded in the allocation of treatment groups.

Procedure

The *S* sat at a horizontal panel with the middle and index finger of each hand resting on a pair of response buttons. In those conditions in which the pairs of buttons were laterally separated, the right hand operated the pair on the right of the panel,

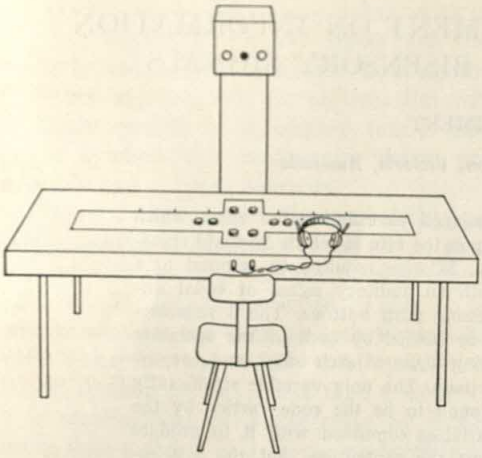


FIG. 1. Layout of signal display and response panel.

TABLE 1

EXPERIMENTAL DESIGN

Code	Digit corresponding				Digit noncorresponding			
	Separation (in inches)							
	0	6	12	24	0	6	12	24
Limb-mode	4	4	4	4	4	4	4	4
Limb-direction	4	4	4	4	4	4	4	4

Note.—Cell entries are number of *S*s used.

and the left hand those on the left. In conditions of zero separation, that is, the buttons being arranged in a rectangle directly in front, the right hand operated the pair farthest from *S*, and the left hand the pair closest to him. Figure 1 illustrates the layout of the response panel.

The operating digits of the two hands carried either corresponding (DC) or noncorresponding (DN) codes. (See Figure 2.) For example, where the limb was coded to the signal's direction of origin, both index fingers responded to the visual component of the signal, one to direction right and the other to direction left, as in Code Condition 3. Alternatively, one digit responded to the visual and the other to the auditory components, as in Condition 4. Similar conditions of digit correspondence applied when the code carried by the limbs was related to the sensory mode of the signal components. In this case, correspondence was defined as both digits carrying the same directional code, as in Condition 1, or different codes, as in Condition 2. The experimental design used is illustrated in Table 1 and the display-control relationships generated by this design in Figure 2. Instructions appropriate to the code condition were read to *S*s from a prepared text. This was followed by a training session consisting of two blocks of 100 signals each, separated by an interval of 3 min. A test series of 60 signals was then presented. The previous study had shown that performance in terms of the information transfer rate (ITR) had continued to improve over 10 test trials of 20 signals each for the 80 *S*s used. The training session employed in this study aimed at obtaining performance measures which were relatively unaffected by gross learning effects.

RESULTS

The dependent variable used for analysis in this study was the ITR achieved by *S*.

The multivariate analysis technique suggested by McGill and summarized by Attneave (1959) was used to calculate the information measures used. ITRs were obtained by dividing the absolute amount of information transmitted by *S* ($T[x;y]$) by the mean reaction time. The obtained ITRs were sub-

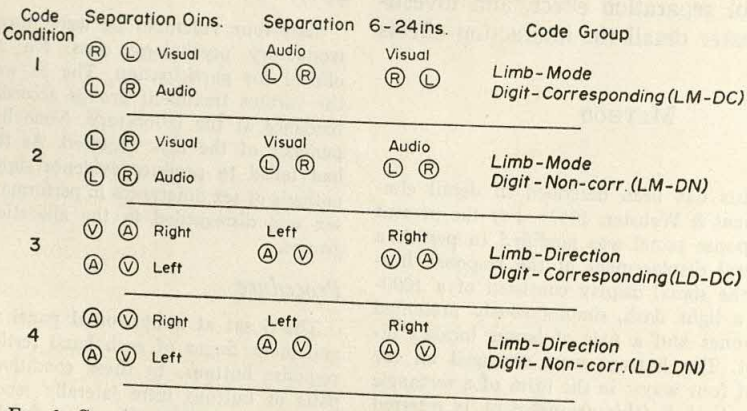


FIG. 2. Control-display arrangements for Limb:Digit:Separation conditions.

TABLE 2

ANALYSIS OF VARIANCE SUMMARY OF THE OBTAINED INFORMATION TRANSFER RATES (BITS/SEC)

Source	*df*	*MS*	*F* ratios
Control separation (A)	3	.16420	1.683
Digit code (B)	1	.26596	2.726
Limb code (C)	1	1.62741	16.679**
A × B	3	.29655	3.039*
A × C	3	.64694	6.626**
B × C	1	.20461	2.097
A × B × C	3	.03925	.402
Error	48	.09757	
Total	63		

Note.—The homogeneity of the treatment variances was tested using Hartley's F_{max} test (Winer, 1962). The obtained *F* ratio of 65 was well below the level required for significance at the .05 level.
* $p < .05$.
** $p < .01$.

jected to an analysis of variance, the results of which are summarized in Table 2.

Examination of Table 2 shows that the only variable making an independent contribution to the total variance is that of the limb code used. The mean ITR obtained where the operating limb responded to the mode of the signal components was 1.17 bits/sec. When the limb responded to the directional components, ITR increased to 1.49 bits/sec. That this effect was not consistent for all treatment groups is shown by the significance of the A × C interaction (Figure 3). The significant feature of this effect is that it is confined exclusively to the displacement of the response buttons from a position on the midplane of *S* to the first position of lateral separation.

To illustrate the effect of control separation under each of the four response codes, treatment means are plotted separately in Figure 4.

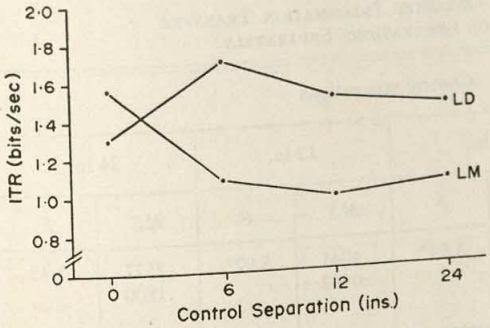


FIG. 3. Interaction of limb code and control separation. (LD = limb direction; LM = limb mode.)

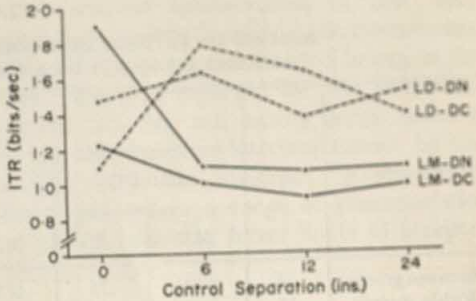


FIG. 4. Interaction of digit code and control separation. (Comparison between limb codes: LD = limb direction; LM = limb mode; DC = digit corresponding; DN = digit noncorresponding.)

As the values plotted in Figure 4 represent the performance of the 16 treatment groups separately, independent tests were made to determine the significance of the differences displayed between limb-digit code groups for each of the lateral separation conditions, and between each of the separation conditions for each limb-digit code group. (See Tables 3 and 4.) Table 3 summarizes the effect upon ITR of laterally separating the response buttons for each of the limb-digit code conditions. The impression conveyed by Figure 4 is supported by the results of these analyses. Only the limb-mode–digit-noncorresponding condition was significantly affected by control separation, and this was restricted to the displacement from 0 to 6 in. separation. Table 4 presents the results of the separate analyses made of the differences between the means obtained by each of the limb-digit groups for the lateral separation conditions. The differences between the limb codes found in the main analysis of variance (Table 2) are reflected in these separate analyses, with the exception that the groups operating under the condition where the controls were separated by 24 in. did not differ significantly in performance.

DISCUSSION

One feature of the results obtained from the present study, which is incongruent with those obtained earlier, relates to the significance of the control separation factor. In contrast to the superiority of centrally located

TABLE 3

ANALYSIS OF VARIANCE SUMMARIES OF OBTAINED INFORMATION TRANSFER RATES (BITS/SEC): FOUR CODING ARRANGEMENTS SEPARATELY

Source	*df*	Control code							
		LM–DC		LM–DN		LD–DC		LD–DN	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Between groups	3	.0791	.75	.6753	16.98*	.3435	1.96	.0484	.69
Within groups	12	.1049		.0397		.1754		.0712	
Total	15								

Note.—LM = limb mode; LD = limb direction; DC = digit corresponding; DN = digit noncorresponding.
* $p < .01$.

controls, which was demonstrated in the earlier study, the present experiment produced no evidence of significant differences in performance which could be related directly to this factor. The explanation of this apparent inconsistency may be approached by consideration of the experimental designs used in the two studies. The previous study employed only two levels of each factor, giving a 2 × 2 × 2 design in which the control separation factor had only two levels, centrally located, and 30 in lateral separation. As one of the aims of the present study was to determine the function which described the influence of control separation on ITR, a 2 × 2 × 4 design was used in which the separation factor had four levels, one of zero separation, and three of lateral separation. The results of the present study replicate, to some extent, those obtained previously in that significant differences in performance were found between some groups operating with central control placement and those operating with controls laterally separated (Figure 4). Further increases in the extent of this displacement had no significant effect upon performance, however. With the display-control relationships used in this study, the effect of control placement upon ITR was found to be of significance for some limb-digit codes, but not for others. Where the limb-digit code was affected by this factor, the significant displacement was from a central location to that where the button pairs were separated by 6 in.

As stated earlier, the aim of this study was to examine in detail the nature of the interactions previously observed between the three response variables examined. Although the question of which, if any, of the display-control arrangements was the "best" was not considered, it is felt that some comments related to this issue would not be inappropriate.

TABLE 4

ANALYSIS OF VARIANCE SUMMARIES OF OBTAINED INFORMATION TRANSFER RATES (BITS/SEC): FOUR CONTROL SEPARATIONS SEPARATELY

Source	*df*	Control separations							
		0 in.		6 in.		12 in.		24 in.	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Between groups	3	.4854	10.65**	.5342	5.54*	.4041	5.92*	.2577	1.43
Within groups	12	.0455		.0964		.0682		.1800	
Total	15								

* $p < .05$.
** $p < .01$.

In the course of analyzing the data from the present experiment, Duncan's New multiple-range Test (Edwards, 1963) was applied to the means of the 16 treatment groups. It was found that the upper eight values plotted in Figure 4 formed a subset of means which did not differ significantly from each other at the .05 level. It is of interest to note that the treatment groups concerned represent the following coding arrangements: limb-direction–digit-noncorresponding = 0 separation; limb-direction–digit-noncorresponding = 6–24 in. separation; limb-direction–digit-corresponding = 6–24 in. separation; limb-mode–digit-noncorresponding = 0 separation. This represents a considerable variety of display-control relationships, and gives emphasis to the concept of flexibility which is often quoted as an attribute of major importance of the human component in man-machine systems.

The results of this experiment are in accord with previous studies concerned with display-control relationships, in that they again demonstrate the effect upon performance of the S-R expectancies which *S* brings to the task. They also highlight the now well-documented fact that satisfactory design features of the man-machine interface cannot be described on a continuum of "good, better, best," but reflect a range of characteristics which has relatively broad limits of acceptability.

REFERENCES

ATTNEAVE, F. *Applications of information theory to psychology*. New York: Holt-Dryden, 1959.

EDWARDS, A. L. *Experimental design in psychological research*. New York: Holt, Rinehart & Winston, 1963.

PERRIMENT, A. D., & WEBSTER, W. R. Display-control relationships with bisensory signals. *Human Factors*, 1967, **9**, 5.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received October 9, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 5, 362–365

EVOLUTIONARY PERFORMANCE, MANAGERIAL ABILITIES, AND CHANGE:

AN EXPLORATORY INVESTIGATION OF ORGANIZATIONS

ERIC P. SANDERS [1]

San Diego State College

In the laboratory, 50 3-man simulated organizations were observed, each operating a model railroad for 1 hr. and each given identical instructions. For each *S* ($N = 150$), data were collected from the Ghiselli Self-Description Inventory, the Gough Adjective Check List, and the performance variable "number of cars per train." The main measurements used were means, standard deviations, and skewness. It was shown that certain properties of an organization's membership can and do affect performance through the evolution of the organization, and that changes of both a task and functional nature continually decrease from the inception to the completion of an organization.

Over the past few years there has been a change in the types of problems which have drawn the attention of investigators in industrial psychology. One of the new areas which has commanded interest may be termed "organizational psychology." As Schein (1965) has indicated, this new field emphasizes the "recognition that organizations are complex systems . . . and human behavior within the organization has to be viewed from the perspective of the entire . . . system [p. 3]."

The present problem is concerned not with the study of individuals comprising groups, but rather with an examination of the total organization's behavior as a complex entity, where the interest is in the exploration of changes in the organization's dynamic "lifespan."

The term "organization" as used in this study will be defined as a social group where the individuals within it are differentiated with respect to their function in connection with a task of achieving some common goal (after Stogdill, 1950).

Organizational evolution is defined as the maturation of the organization from its inception to its completion. Ghiselli (1966) and Ghiselli and Lodahl (1958) have studied properties of groups in relation to performance, and some work has been done in viewing organizational properties (e.g., Ghiselli, 1963; Smelser, 1961). However, the relationship between organizational properties and performance over time has not been investigated and is the concern of the present study.

The other part of this investigator's consideration will be focused upon "organizational stability." As the organization develops it tends to become fixed in its procedures. This is implicitly stated in Weber's (1946) classic bureaucratic theory and other theories in small group research, but it has not, to this researcher's knowledge, been examined in an experimental setting.

The two major problems, therefore, upon which attention is focused in this investigation are:

(*a*) organizational evolution and the properties of the organization's membership which affect performance during evolution, and

(*b*) organizational stability—whether or not an organization maintains its original functions or whether they are changed and modified by the members over time.

METHOD

In order to examine the problems raised above, a complex model railroad task was developed. This type of apparatus, but on a much simpler scale, was used before by Ghiselli and Lodahl (1958), Ghiselli (1966), and Smelser (1961) for studying group performance. The railroad is a useful task for the present purpose because it is inherently interesting and motivating for the young men used as *S*s.

Fifty three-man organizations worked on the actual

[1] A study done in partial fulfillment of the requirements for a doctoral degree at the University of California, Berkeley. The author wishes to thank Edwin E. Ghiselli, University of California, Berkeley, for his judicious help and constructive criticism which helped make this study possible.

task for a total time of 1 hr. Participants were instructed that it was a cooperative task and each group was given identical instructions. None of the members had worked together prior to the experimental situation. The task was subdivided into thirds: the first third (Trials 1–4); the second third (Trials 5–8); and the last third (Trials 9–12). Each trial was 5 min. in length.

After completing the task session, *Ss* were given the Ghiselli Self-Description Inventory (SDI) and the Gough Adjective Check List (ACL). The former gives an indication of the extent to which individuals possess certain managerial abilities. The six scales of the Ghiselli SDI relevant to this investigation were titled: Intelligence, Initiative, Supervisory Ability, Perceived Occupational Level, Decision-Making Approach, and Maturity. The Gough ACL is used to describe certain psychological attributes of a person. The selected scales used were: Achievement, Dominance, and Aggressiveness. The average reliability for both the Ghiselli SDI and the Gough ACL was .80. For each simulated organization the mean, range, and skewness of scores were determined for each of the properties measured by the SDI and ACL.

The reliability of performance was determined by the odd-even correlation method. As a measure of "efficiency" of performance (rather than just sheer amount), "cars per train" was used. The uncorrected reliability of this measure was .70.

In order to ascertain whether or not changes took place, specific job changes and changes in functional procedures were recorded.

RESULTS

The data indicate that as organizations evolve performance improves considerably. The mean number of "cars per train" (efficiency) increased from 3.80 to 4.40 from Stage I to Stage III, respectively, the difference being significant at the .01 level.

Although one might expect that, as the intelligence and initiative of the organization members, on the average, increase, the better would be the performance at the inception of the organization, Table 1 shows this relationship does not appear to be the case for average intelligence, though it does seem to be so for initiative. That is, one's expectation would be for a higher correlational relationship at the early stages of evolution and a lower relationship at the later stages. The data do not show this to be the case.

TABLE 1

CORRELATION BETWEEN AVERAGE INTELLIGENCE, AVERAGE INITIATIVE, AND AVERAGE MANAGERIAL ABILITIES AND PERFORMANCE THROUGH THE INITIAL AND FINAL STAGES OF ORGANIZATION EVOLUTION

Properties	Stage I	Stage III	Difference
Average intelligence	.08	.22	*ns*
Average initiative	.15	−.12	*ns*
Average managerial ability	.06	−.21	2.40*

* $p = .05$.

TABLE 2

MEAN DIFFERENCES BETWEEN HIGH AND LOW DMA DIFFERENTIATION AND PERFORMANCE DURING EARLY AND LATER STAGES OF ORGANIZATION EVOLUTION

Stages	Low DMA skew means	High DMA skew means	Difference
I	46.6	50.0	.93
III	46.0	51.0	1.32*

* $p > .10$.

The various "managerial abilities" were formed into a composite "managerial abilities variable" following Ghiselli's (1964) formula, and then correlated with performance during the initial and final stages of organization evolution. The data indicate that fewer managerial abilities are manifested, on the average, by the organization membership as performance improves over time.

If there were one person who was more highly differentiated with regard to his decision-making approach (DMA) than the other members of the organization, the better would be the organization's later performance (Table 2). The groups that distinguished the strong decision maker did perform better than those that did not. This is most readily seen in the differences in performance during the final evolutionary stage (Stage III).

It appears to be quite significantly established, as shown in Table 3, that there is

TABLE 3

DIFFERENCES IN MEAN NUMBER OF ORGANIZATION CHANGES DURING THE STAGES OF ORGANIZATION EVOLUTION

Stages	II	III	I & II	Mean changes
I	2.70*	4.00*		.72
II		3.00*		.46
III			5.37*	.16

* $p > .005$.

greater stability (fewer organization changes) at the later stages of organization evolution than at the earlier stages. The number of changes dropped off some 80% from the first to the last stage.

DISCUSSION

Generally speaking, performance throughout the "life" of the organization continually improves. This supports the results found by Ghiselli (1966) in that the "learning" of total organizations appears to follow a similar pattern to that of individuals—practice improving performance with the improvement shown in the characteristic negative acceleration. Since this "learning" or performance was measured in terms of a single organizational score (number of cars per train), rather than just the sum of aggregate member scores, it can be termed organizational learning.

The mean intelligence level of the groups forming the organizations is unrelated to performance whether early or later in the organization's history. This agrees with Ghiselli's (1966) results which indicate that organization performance is not necessarily related to the mean scores for the organization. Since intelligence is defined here as being a wide domain of mental abilities rather than any particular traits, it is possible that performance during the initial stages of organizational evolution may not be related to the mean intelligence of the organization members, but may be accounted for only by those certain properties associated with particular behavior at this evolutionary stage (e.g., ability to deal with ambiguity, etc.), which may be masked by using a mean score measure, the implication being that coordination among members of an organization may be impeded in cooperative situations by good problem solvers (implied high intelligence).

Initiative, on the other hand, presents another picture. Initiative is defined by Ghiselli (1956) as "involving the beginnings of actions and the capacity to note or discover new means of goal achievement [p. 311]." This implies acting independently of others and the ability to develop some sort of structure from a nonstructured situation. If an organization has members who score higher, on the average, on initiative, the better will be the performance during the initial stages since these individuals will be better able to structure the situation. Such appears to be the case with the groups studied.

It seems that if an organization differentiates a single member who is a strong decision maker (high positive skewness), while the remainder of the members are relatively homogeneous, then the later performance of that organization tends to improve. Since there is a relationship between those members who score high on the DMA skewness scale and high in initiative ($r = .24$, $p > .10$) and since initative is significantly related to certain managerial abilities (e.g., supervisory ability, $r_{is} = .48$; perceived occupational level, $r_{ip} = .59$; achievement, $r_{ia} = .39$; and dominance, $r_{id} = .45$; all significant at the .01 level), one would expect that an uncontested strong decision maker in the organization would not only be able to influence other member's performance but also, resulting from his self-confidence and strong initiative, maintain its upward direction through the organization's evolution. Generalizations to all types of organizations, however, must be made with caution since the results are based on loosely structured task groups which allow flexibility of interaction as well as complete freedom in sharing in the decision-making process. With a strong hierarchically structured "formal organization," these results might show a different type of relationship. However, in this situation, it is evident that with a highly differentiated decision maker, the organization will perform better later. These results coincide with some of the classic bureaucratic theories of Weber (1946) and the studies by Ghiselli and Lodahl (1958) and Ghiselli (1966).

The classic bureaucratic assumption regarding organization stability is tenable, according to these data. This indicates that at the inception of an organization there tends to be an unstructured, ambiguous situation wherein the members are just trying to understand their particular tasks, trying to fit their tasks into the total pattern of organizational functioning, and looking for some unifying direction. Resulting from this situation, more changes occur in the tasks, procedures, and structure during the early stages of evolu-

tion, which is to say that the stability of organizations increases throughout organizational evolution.

REFERENCES

GHISELLI, E. E. Correlates of initiative. *Personnel Psychology*, 1956, 9, 311–320.

GHISELLI, E. E. Managerial talent. *American Psychologist*, 1963, 18, 631–642.

GHISELLI, E. E. *Theory of psychological measurement*. New York: McGraw-Hill, 1964.

GHISELLI, E. E. Psychological properties of groups and group learning. *Psychological Reports*, 1966, 19, 17–18.

GHISELLI, E. E., & LODAHL, T. M. Patterns of managerial traits and group effectiveness. *Journal of Abnormal and Social Psychology*, 1958, 57, 61–66.

SCHEIN, E. H. *Organizational psychology*. Englewood Cliffs, N. J.: Prentice-Hall, 1965.

SMELSER, W. T. Dominance as a factor in achievement and perception in cooperative problem solving interaction. *Journal of Abnormal and Social Psychology*, 1961, 62, 535–542.

STOGDILL, R. M. Leadership, membership and organization. *Psychological Bulletin*, 1950, 47, 1–14.

WEBER, M. *Essays in sociology*. New York: Oxford University Press, 1946.

(Received October 9, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 5, 366–371

TESTING TRADITIONAL AND TWO-FACTOR HYPOTHESES CONCERNING JOB SATISFACTION

GEORGE B. GRAEN [1]

University of Illinois

Traditional and 2-factor theories make contradicting hypotheses concerning the relationships between satisfaction with job factors and overall job satisfaction. Traditional theory predicts essentially linear relationships, whereas 2-factor theory predicts particular nonlinear relationships. *S*s were 167 male and 152 female office workers employed by a single corporation. Data collected included satisfaction with work itself, promotion, and pay as measured by the scales of the Job Description Index and overall job satisfaction as measured by the General Motors Faces Scale. Data were analyzed by analysis of variance. Results indicated not only that 2-factor predictions were not confirmed on 4 out of 6 relationships but also that 3 of these 4 misses were opposite that predicted. Traditional predictions generally were supported.

The two-factor theory of job satisfaction (Herzberg, 1966; Herzberg, Mausner, & Snyderman, 1959) hypothesizes two sets of job variables—job-content or "satisfier" variables and job-context or "dissatisfier" variables. Moreover, these two sets of variables (job-content and job-context) are assumed to contribute to overall job satisfaction in different manners. Job-content variables are assumed to account primarily for variance on the satisfaction portion of the continuum of job satisfaction–dissatisfaction. According to two-factor theory, when the positive aspects of job-content variables, such as work itself and advancement, are present on the job, they can increase job satisfaction from indifference to lust for the job. However, when the negative aspects of these variables are present, they can do little or nothing to promote dissatisfaction. In contrast, when the positive aspects of job-context variables, such as salary, are present, they cannot, according to the theory, increase satisfaction much above indifference. However, when their negative counterparts are present, they can increase dissatisfaction from indifference to scorn for the job. Thus, two-factor theory predicts nonlinear relationships between satisfaction with job variables and overall job satisfaction.

As an alternative formulation, the traditional theory of job satisfaction (Ewen, Smith, Hulin, & Locke, 1966; Graen, 1966) hypothesizes that if job variables are categorized according to the two-factor classification, both job-content and job-context variables account for variance over the entire satisfaction–dissatisfaction continuum. Moreover, job-content variables are more potent contributors to satisfaction–dissatisfaction than are job-context variables. Traditional theory hypothesizes that if the presence of a variable contributes to job satisfaction, the absence of that variable will contribute to dissatisfaction and vice versa. Thus, traditional theory predicts essentially linear relationships between satisfaction with job variables and overall job satisfaction. Viewed in this way, two-factor and traditional theories specify different functions between job variables and overall job satisfaction. Two-factor theory predicts nonlinear relationships; traditional theory predicts essentially linear relationships.

Two recent articles (Ewen et al., 1966; Graen, 1966) have presented evidence indicating the two-factor hypotheses do not hold for *work itself, promotion,* or *pay*. These investigators found that dissatisfaction with the job-content variables (work itself and promotion) contributed to overall dissatisfaction and that satisfaction with the job-context variable (pay) contributed to overall satisfaction. Taken as a whole, these results clearly supported traditional theory at the expense of two-factor theory.

These results of Ewen and Graen using male industrial workers so clearly contradicted

[1] The author wishes to thank Marvin D. Dunnette and Charles L. Hulin for their cooperation and support.

two-factor predictions that further research is needed to determine their generality. The purpose of the present study was to test the generality of the Ewen and Graen findings by applying their methods to samples of male and female office workers. As in the Ewen and Graen studies, two major hypotheses were tested on which two-factor and traditional theories conflict. These hypotheses are stated below.

Hypothesis 1. If one group of employees is *dissatisfied,* a second group *neutral,* and a third group *satisfied* with regard to a job-content (satisfier) variable, two-factor theory predicts equal overall satisfaction for the first and second groups and higher satisfaction than the other two groups for the third. In contrast, traditional theory predicts an increasing linear trend in satisfaction from the first to the third group.

Hypothesis 2. If one group of employees is *dissatisfied,* a second group *neutral,* and a third group *satisfied* with regard to a job-context (dissatisfier) variable, two-factor theory predicts equal overall satisfaction for the second and third groups and lower satisfaction than the other two groups for the first. In contrast, traditional theory again predicts an increasing linear trend in satisfaction from the first to the third group.

In addition to these two major hypotheses, the Ewen and Graen studies generated two other hypotheses that will be tested in the present study. These two hypotheses are stated below.

Hypothesis 3. Job-content variables are more strongly related to overall satisfaction–dissatisfaction than are job-context variables.

Hypothesis 4. The relationships between job-context variables and overall satisfaction–dissatisfaction depend on the level of satisfaction with the job-content variables.

METHOD

The *S*s were home-office employees of one corporation. All employees were surveyed as part of a job-attitudes study conducted by the company. A mail questionnaire assuring anonymity to the respondents was administered through departmental representatives. A gratifyingly high rate (83%) of return was obtained and all questionnaires were usable. The returns included 167 questionnaires from males and 152 from females.

Among other instruments included in the questionnaire were the Job Description Index (JDI) and the General Motors Faces scale. The JDI developed at Cornell is an adjective checklist dealing with five job variables, namely, work itself, promotion, pay, supervision, and co-workers. The JDI has been shown to possess acceptable convergent and discriminant validity (Hulin, Smith, Kendall, & Locke, 1963; Kendall, Smith, Hulin, & Locke, 1963; Locke, Smith, Hulin, & Kendall, 1963; Locke, Smith, Kendall, Hulin, & Miller, 1964). This study employed only the work itself, promotion, and pay scales of the JDI. The General Motors Faces Scale (Kunin, 1955) was used as the measure of overall job satisfaction. For details of the data-collection procedure, see Hulin and Smith (1967).

In the analyses of Ewen and Graen, the variables of work itself and promotion were analyzed as a composite variable. In the present study, work itself and promotion were analyzed separately. Analyses employed a weighted means, two-way analysis of variance to test the relationships between the JDI variables of work itself, promotion, and pay and overall job satisfaction measured by the General Motors Faces Scale. These analyses were performed separately for males and females. Each of the three JDI variables (work itself, promotion, and pay) was trichotomized into dissatisfied, neutral, and satisfied with regard to the job variable. Cutting points on these JDI variables were supplied by Ewen (1965).

In all ANOVAs, the first dimension was a job-content variable (either work itself or promotion), and the second was the job-context variable (pay). Each dimension had three levels, namely, dissatisfied, neutral, and satisfied. The criterion measure in all ANOVAs was overall job satisfaction as measured by the General Motors Faces Scale. In addition, a priori contrasts were tested on the main effects (Winer, 1962). For each dimension, one contrast (A_1 and B_1) tested the difference on overall satisfaction between the group *dissatisfied* and the group *neutral* with regard to the JDI dimension. A second contrast (A_2 and B_2) tested the difference on overall satisfaction between the group *neutral* and the group *satisfied* with regard to the JDI dimension. Omega-squared (Hays, 1963) was employed to measure the strength of relationships between JDI dimensions and overall job satisfaction. Finally, the trend in means on overall satisfaction for groups dissatisfied, neutral, and satisfied with regard to each JDI variable was tested for curvilinearity (McNemar, 1962).

RESULTS

Results of the ANOVAs for male employees are shown in Tables 1 and 2. According to Tables 1 and 2, work itself was related to high and not low overall satisfaction; promotion was related to low and not high satisfaction; pay was related to high and not low satisfaction. Thus, work itself was predicted and promotion was not predicted in Hypothe-

TABLE 1

ANALYSIS OF VARIANCE FOR A PRIORI CONTRASTS AND OMEGA-SQUARED FOR STANDARD EFFECTS OF THE SATISFIER (WORK) AND THE DISSATISFIER (PAY) DIMENSIONS ON OVERALL JOB SATISFACTION FOR MALES

Source of variation	*df*	*MS*	*F*	Omega²
Satisfier (work)	2	31.88	13.23**	.12
A_1: S_D versus S_N	1	6.89	2.86	
A_2: S_N versus S_S	1	23.04	9.56**	
Dissatisfier (pay)	2	8.70	3.61*	.03
B_1: D_D versus D_N	1	.10	.04	
B_2: D_N versus D_S	1	11.06	4.59*	
Interaction	4	5.58	2.23	.03
Error	158	2.41		

Note.—S and D represent the satisfier and dissatisfier factors, respectively. Subscripts D, N, and S represent the dissatisfied, neutral, and satisfied levels, respectively.
* $p \leq .05$.
** $p \leq .01$.

sis 1 by the two-factor theory. Moreover, promotion not only failed to support this two-factor hypothesis, but actually demonstrated a relationship the reverse of that predicted. Two-factor theory predicted, in Hypothesis 1, that content variables should show significant differences, if any, between neutral and satisfied groups and *not* between dissatisfied and neutral groups. Although this was the case for work itself, the opposite was the case for promotion. In Hypothesis 2, two-factor theory predicted that context variables should show significant differences, if any, between dissatisfied and neutral groups and *not* between neutral and satisfied groups. In contradiction to this hypothesis, pay showed a relationship opposite that predicted by two-factor theory. Overall, these results for male employees confirm two-factor predictions for only a single content variable (work itself) and strongly disconfirm two-factor predictions for both the second content variable (promotion) and the context variable (pay).

Results of the ANOVAs for female employees are presented in Tables 3 and 4. As shown in Tables 3 and 4, work itself was related to both high and low overall satisfaction, and both promotion and pay were related to high but not low satisfaction. In the case of females, work itself was not predicted and promotion was predicted in Hypothesis 1 by two-factor theory. In agreement with the results for males, pay showed a relationship opposite that predicted in Hypothesis 2 by two-factor theory. These results for females were similar to those for males with the exception that promotion was predicted and work itself was not predicted. Therefore, two-factor predictions fared no better in either sample: Hypothesis 1 was supported by one variable and not supported by the other variable; Hypothesis 2 was not supported.

Results on the predictions of traditional theory are shown in Table 5. According to Table 5, the means on overall satisfaction

TABLE 2

ANALYSIS OF VARIANCE FOR A PRIORI CONTRASTS AND OMEGA-SQUARED FOR STANDARD EFFECTS OF THE SATISFIER (PROMOTION) AND THE DISSATISFIER (PAY) DIMENSIONS ON OVERALL JOB SATISFACTION FOR MALES

Source of variation	*df*	*MS*	*F*	Omega²
Satisfier (promotion)	2	33.08	13.48**	.12
A_1: S_D versus S_N	1	12.38	5.05*	
A_2: S_N versus S_S	1	6.04	2.47	
Dissatisfier (pay)	2	13.94	5.68**	.05
B_1: D_D versus D_N	1	2.55	1.04	
B_2: D_N versus D_S	1	9.08	3.71	
Interaction	4	3.16	1.29	.01
Error	158	2.45		

Note.—S and D represent the satisfier and dissatisfier factors, respectively. Subscripts D, N, and S represent the dissatified, neutral, and satisfied levels, respectively.
* $p \leq .05$.
** $p \leq .01$.

TABLE 3

ANALYSIS OF VARIANCE FOR A PRIORI CONTRASTS AND OMEGA-SQUARED FOR STANDARD EFFECTS OF THE SATISFIER (WORK) AND THE DISSATISFIER (PAY) DIMENSIONS ON OVERALL JOB SATISFACTION FOR FEMALES

Source of variation	*df*	*MS*	*F*	Omega²
Satisfier (work)	2	91.50	34.81*	.27
A_1: S_D versus S_N	1	58.06	22.08*	
A_2: S_N versus S_S	1	25.05	9.52*	
Dissatisfier (pay)	2	19.78	7.52*	.05
B_1: D_D versus D_N	1	.67	.25	
B_2: D_N versus D_S	1	31.57	12.00*	
Interaction	4	1.38	.52	.00
Error	143	2.63		

Note.—S and D represent the satisfier and dissatisfier factors, respectively. Subscripts D, N, and S represent the dissatisfied, neutral, and satisfied levels, respectively.
* $p \leq .01$.

TABLE 4

ANALYSIS OF VARIANCE FOR A PRIORI CONTRASTS AND OMEGA-SQUARED FOR STANDARD EFFECTS OF THE SATISFIER (PROMOTION) AND THE DISSATISFIER (PAY) DIMENSIONS ON OVERALL JOB SATISFACTION FOR FEMALES

Source of variation	*df*	*MS*	*F*	Omega²
Satisfier (promotion)	2	36.20	10.73**	.10
A_1: S_D versus S_N	1	4.00	1.19	
A_2: S_N versus S_S	1	20.91	6.20*	
Dissatisfier (pay)	2	29.50	8.74**	.08
B_1: D_D versus D_N	1	.45	.13	
B_2: D_N versus D_S	1	33.36	9.90**	
Interaction	4	2.34	.69	.00
Error	143	3.37		

Note.—S and D represent the satisfier and dissatisfier factors, respectively. Subscripts D, N, and S represent the dissatisfied, neutral, and satisfied levels, respectively.
* $p \leq .05$.
** $p \leq .01$.

TABLE 5

MEANS ON OVERALL JOB SATISFACTION FOR THE GROUPS DISSATISFIED, NEUTRAL, AND SATISFIED WITH REGARD TO WORK ITSELF, PAY, AND PROMOTION

Group	Work		Pay		Promotion	
	n	*M*	*n*	*M*	*n*	*M*
Males						
Dissatisfied	12	6.17	21	7.19	44	7.14
Neutral	18	7.17	29	7.79	33	8.12
Satisfied	137	8.47	117	8.43	90	8.68
Females						
Dissatisfied	19	5.21*	18	6.89	61	7.36
Neutral	26	7.73	22	7.09	27	8.11
Satisfied	107	8.89	112	8.67	64	9.11

Note.—Standard deviations on overall satisfaction were 1.73 and 2.07 for males and females, respectively.
* Significantly nonlinear at $p \leq .01$.

demonstrated an increasing monotonic trend for groups dissatisfied, neutral, and satisfied with the job variables. This relationship, predicted by traditional theory, was shown in both samples by work itself, promotion, and pay. Of the six relationships tested for curvilinearity, only that between work itself and satisfaction for females was significantly curvilinear. However, the trend in means on satisfaction for this single curvilinear relationship was opposite that predicted by two-factor theory. Although two-factor theory predicted no difference in satisfaction between the groups dissatisfied and neutral with regard to work itself, the largest difference was between these very groups. Thus, predictions made by traditional theory in Hypotheses 1 and 2 were supported by these data. Moreover, these data clearly are not compatible with two-factor predictions of nonlinear relationships between satisfaction with job variables and overall satisfaction.

The results relevant to Hypothesis 3 are the measures of strength of association (Omega-squared). These results strongly support the hypothesis that content variables (work itself and promotion) are more highly related to overall satisfaction–dissatisfaction than context variables (pay). In the male sample, work itself and promotion accounted for 12% of the total variance in overall satisfaction, and pay accounted for 3% and 5%. Similarly, in the female sample, work itself accounted for 27% of the total variance in satisfaction, promotion accounted for 10%, and pay for 5% and 8%. Further evidence of the large differences in strength of relationship can be seen in the differences between means shown in Table 5.

Results on Hypothesis 4 are the ANOVA tests on interactions. According to Tables 1 through 4, none of the interactions between content and context dimensions was significant. Therefore, the hypothesis that the functioning of context variables depends on the level of satisfaction with the content variables was not supported by these data.

DISCUSSION

The results of this study, taken as a whole, must be viewed as damaging to the two-factor theory. Although failure by two-factor theory to predict four out of six relationships seems discouraging enough, three of these four misses actually were *opposite* that predicted. In addition, the trends of the means on overall satisfaction for the groups dissatisfied, neutral, and satisfied with the job variables were increasing monotonic and essentially linear, rather than nonlinear as predicted by two-factor theory. In contrast, hypotheses of traditional theory generally were supported in this study. The prediction of increasing

linear trends in satisfaction for groups dissatisfied, neutral, and satisfied with regard to job variables was supported for both the content variables of work itself and promotion and the context variable of pay. These findings held for both males and females with a single exception.

This study also supports the conclusions of Friedlander (1964), Wernimont (1966), Ewen et al. (1966), and Graen (1966) that content (intrinsic) variables are more strongly related to overall satisfaction–dissatisfaction than are context (extrinsic) variables. In the previous study by Graen (1966) the content variable accounted for 18% and the context variable 2% of the total variance in overall satisfaction. In the present study, the content variables accounted for between 10% and 27% and the context variable accounted for between 3% and 8%.

In agreement with Graen (1966), this study failed to yield a significant interaction between content and context variables. Thus, the hypothesis by Ewen et al. (1966) that the functioning of extrinsic variables depends on the level of satisfaction with the intrinsic variables finds no support in this study.

One major difference in results appeared between this replication study and the original studies by Ewen and Graen. The sample employed by Ewen and Graen consisted of male employees selected to represent the population of industrial concerns employing 50 or more persons (Kendall, 1963). In those studies, the content variable (a composite of work itself and promotion) and the context variable (pay) contributed significantly to both high and low overall satisfaction. In the male sample of the present study, work itself contributed only to high satisfaction, promotion only to low satisfaction, and pay only to high satisfaction. The main difference between these two sets of results was that pay failed to contribute to low satisfaction in the present study. This difference probably was due to the situational variables associated with a single company.

Even within the same company, this study found large sex differences. In the sample of female employees, work itself contributed to both high and low overall satisfaction, and promotion and pay only to high satisfaction. In contrast, for males working in the same location, promotion contributed only to low satisfaction, and work itself and pay only to high satisfaction.

This study illustrates once again that the major hypotheses of the two-factor theory, in their existing form, cannot predict results produced by the present method. Although two-factor hypotheses summarize the results produced by the story-telling method very well, they appear useless in predicting results of studies employing other methods. For documentation on this point, see Dunnette, Campbell, and Hakel (1967). Unless the two-factor theory can be modified so as to account for the disconfirming results obtained when different methods have been employed it should be regarded as of limited usefulness in the attempt to understand work motivation.

REFERENCES

DUNNETTE, M. D., CAMPBELL, J. P., & HAKEL, M. D. Factors contributing to job satisfaction and job dissatisfaction in six occupational groups. *Journal of Organizational Behavior and Human Performance,* 1967, **2**, 143–174.

EWEN, R. B. Weighting components of job satisfaction. Unpublished doctoral dissertation, University of Illinois, 1965.

EWEN, R. B., SMITH, P. C., HULIN, C. L., & LOCKE, E. A. An empirical test of the Herzberg two-factor theory. *Journal of Applied Psychology,* 1966, **50**, 544–550.

FRIEDLANDER, F. Job characteristics as satisfiers and dissatisfiers. *Journal of Applied Psychology,* 1964, **48**, 388–392.

GRAEN, G. B. Addendum to "An empirical test of the Herzberg two-factor theory." *Journal of Applied Psychology,* 1966, **50**, 551–555.

HAYS, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

HERZBERG, F., MAUSNER, B., & SNYDERMAN, B. B. *The motivation to work.* (2nd ed.) New York: Wiley, 1959.

HERZBERG, F. *Work and the nature of man.* New York: World, 1966.

HULIN, C. L., & SMITH, P. A. An empirical investigation of two implications of the two-factor theory of job satisfaction. *Journal of Applied Psychology,* 1967, **51**, 396–402.

HULIN, C. L., SMITH, P. C., KENDALL, L. M., & LOCKE, E. Cornell studies of job satisfaction: II. Model and method of measuring job satisfaction. Cornell University, 1963. (Mimeo)

KENDALL, L. M. Canonical analysis of job satisfaction and behavioral, personal background, and situational data. Unpublished doctoral dissertation, Cornell University, 1963.

KENDALL, L. M., SMITH, P. C., HULIN, C. L., & LOCKE, E. A. Cornell studies of job satisfaction: IV. The relative validity of the job description index and other methods of measurement of job satisfaction. Cornell University, 1963. (Mimeo)

KUNIN, T. The construction of a new type of attitude measure. *Personnel Psychology*, 1955, **8**, 65–66.

LOCKE, E. A., SMITH, P. C. HULIN, C. L., & KENDALL, L. M. Cornell studies of job satisfaction: V. Scale characteristics of the job description index. Cornell University, 1963. (Mimeo)

LOCKE, E. A., SMITH, P. C., KENDALL, L. M., HULIN, C. L., & MILLER, A. M. Convergent and discriminant validity for areas and methods of rating job satisfaction. *Journal of Applied Psychology*, 1964, **48**, 313–319.

MCNEMAR, Q. *Psychological statistics*. (3rd ed.) New York: Wiley, 1962.

WERNIMONT, P. F. Intrinsic and extrinsic factors in job satisfaction. *Journal of Applied Psychology*, 1966, **50**, 41–58.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received October 18, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 5, 372–376

SIGNS, SAMPLES, AND CRITERIA

PAUL F. WERNIMONT AND JOHN P. CAMPBELL

3M Company, St. Paul, Minnesota *University of Minnesota*

Proposes a new emphasis in the prediction of future job behavior. The essence of the suggested procedure is the establishment of consistencies between relevant dimensions of job-behavior and preemployment-behavior samples obtained from real or simulated situations. If samples instead of signs are employed, a number of prediction and measurement problems seem to be alleviated or at least confronted more directly. An emerging technology of behavior sampling and measurement would appear to put these goals within reach.

Many writers (e.g., Dunnette, 1963; Ghiselli & Haire, 1960; Guion, 1965; Wallace, 1965) have expressed concern about the difficulties encountered in trying to predict job performance, and in establishing the validity of tests for this purpose. In general, their misgivings center around the low validities obtained and misapplications of the so-called "classic validity model." To help ameliorate these difficulties it is proposed here that the concept of validity be altered as it is now applied to predictive and concurrent situations and introduce the notion of "behavioral consistency." By consistency of behavior is meant little more than that familiar bit of conventional wisdom, "The best indicator of future performance is past performance." Surprisingly few data seem to exist to either support or refute this generalization. It deserves considerably more attention.

Some History

It is perhaps not too difficult to trace the steps by which applied psychologists arrived at their present situation. During both World War I and World War II general intelligence and aptitude tests were effectively applied to military personnel problems. Largely as the result of these successes, the techniques developed in the armed services were transported to the industrial situation and applied to the personnel problems of the business organization. From a concentration on global measures of mental ability, validation efforts branched out to include measures of specific aptitudes, interests, and personality dimensions. The process is perhaps most clearly illustrated by the efforts of the United States Employment Service to validate the General Aptitude Test Battery across a wide range of jobs and occupations. In general, testing seemed to be a quick, economical, and easy way of obtaining useful information which removed the necessity for putting an individual on the job and observing his performance over a trial period.

It was in the context of the above efforts that an unfortunate marriage occurred, namely, the union of the classic validity model with the use of tests as signs, or indicators, of predispositions to behave in certain ways (Cronbach, 1960, p. 457), rather than as samples of the characteristic behavior of individuals. An all too frequent procedure was to feed as many signs as possible into the classic validity framework in hopes that the model itself would somehow uncover something useful. The argument here is that it will be much more fruitful to focus on meaningful samples of behavior, rather than signs of predispositions, as predictors of later performance.

The Consistency Model

To further illustrate the point, consider a hypothetical prediction situation in which the following five measures are available:

1. Scores on a mental ability test;
2. School grade-point average (GPA);
3. Job-performance criterion at Time 1;
4. Job-performance criterion at Time 2;
5. Job-performance criterion at Time 3.

Obviously, a number of prediction opportunities are possible. Test scores could be correlated with GPA; school achievement could be correlated with first-year job success; or the test scores and GPA could be combined

in some fashion and the composite used to predict first-, second-, or third-year job performance. All of these correlations would be labeled validity coefficients and all would conform to the classic validity model. It is less clear what label should be attached to the correlation between two different measures of job performance. Few would call it validity; many would probably refer to it as reliability. There seems to be a tendency among applied psychologists to withhold the term validity from correlations between measures of essentially the same behavior, even if they were obtained at two different points in time. That is, the subtleties of the concept of reliability and the ingredients of the classic validity model seem to have ingrained the notion that validity is a correlation between a predictor and a criterion and the two should somehow be dissimilar.

However, each of the 10 correlations that one could compute from the above situation represents the degree of common variation between the two variables, given the appropriateness of the linear correlation model. After all, that is what correlation is all about. In this sense there is no logical reason for saying that some of the coefficients represent validity and others reliability, although there certainly may be in other contexts. An implicit or explicit insistence on the predictor being "different" seems self-defeating. Rather one should really be trying to obtain measures that are as similar to the criterion or criteria as possible. This notion appears to be at least implicit in much of the work on prediction with biographical data where many of the items represent an attempt to assess previous achievement on similar types of activities. Behavior sampling is also the basis on which simulation exercises are built for use in managerial assessment programs.

At this point it should be emphasized that for the consistency notion to be consistent, the measures to be predicted must also be measures of behavior. For example, it would be something less than consistent to use a behavior sample to predict such criteria as salary progression, organizational level achieved, or subunit production. The individual does not always have substantial control over such variables, and, even with the more obvious biasing influences accounted for, they place a ceiling on the maximum predictive efficiency to be expected. Furthermore, they are several steps removed from actual job behavior. In this respect, the authors are very much in accord with Dunnette (1966) who argues strongly for the measurement of observable job behavior in terms of its effect on meaningful dimensions of performance effectiveness. A recently developed method for accomplishing this aim is the behavior retranslation technique of Smith and Kendall (1964). The applied psychologist should reaffirm his mandate and return to the measurement of behavior. Only then will one learn by what means, and to what extent, an individual has influenced his rate of promotion, salary increases, or work group's production.

In general terms, what might the selection or prediction procedure look like if one tried to apply a consistency model? First, a comprehensive study of the job would be made. The results of this effort would be in the form of dimensions of job performance well defined by a broad range of specific behavior incidents which in turn have been scaled with respect to their "criticalness" for effective or ineffective performance.

Next, a thorough search of each applicant's previous work experience and educational history would be carried out to determine if any of the relevant behaviors or outcomes have been required of him or have been exhibited in the past. Items and rating methods would be developed to facilitate judging the frequency of such behaviors, the intensity with which they were manifested, the similarity of their context to the job situation, and the likelihood that they will show up again. These judgments can then be related to similar judgments concerning significant and consistent aspects of an individual's job behavior.

Such a procedure places considerable emphasis on background data and is similar in form to the "selection by objectives" concept of Odiorne and Miller (1966). However, the aim is to be considerably more systematic and to focus on job behavior and not summary "objectives."

After the analysis of background data it might be found that the required job be-

haviors have not been a part of the applicant's past repertoire and it would be necessary to look for the likelihood of that job behavior in a variety of work-sample tests or simulation exercises. A number of such behavior measures are already being used in various management assessment programs.

Finally, individual performance measures of psychological variables would be given wider use where appropriate. For example, the Wechsler Adult Intelligence Scale (Wechsler, 1955) might be used to assess certain cognitive functions. Notice that such a measure is a step closer to actual performance sampling than are the usual kinds of group intelligence tests.

How does the above procedure compare to conventional practice? The authors hope they are not beating at a straw man if the usual selection procedure is described as follows. First, a thorough job analysis is made to discover the types of skills and abilities necessary for effective performance. This is similar to the consistency approach except that the objective seems to be to jump very quickly to a generalized statement of skills and abilities rather than remaining on the behavioral level. The conventional approach next entails a search for possible predictors to try out against possible criteria. Based on knowledge of the personnel selection and individual differences literature, personal experience, and "best guesses," some decisions are made concerning what predictors to include in the initial battery. It is the authors' contention that the classic validity model has forced an undue amount of attention on test and inventory measures at this stage. Witness the large amount of space devoted to a discussion of "test validation" in most books dealing with the selection problem. Again, signs seem to take precedence over samples. Lastly, one or more criterion measures are chosen. Too often the choice seems to be made with little reference to the previous job analysis and is based on a consideration of "objectivity" and relevance to the "ultimate" criterion. Unfortunately, even a slight misuse of these considerations can lead to criteria which are poorly understood. In contrast, working within the framework of a consistency model requires consideration of dimensions of actual job behavior.

It might be added that the above characterization of the conventional approach is meant to be somewhat idealized. Certain departures from the ideal might reinforce the use of signs to an even greater extent. For example, there is always the clear and present danger that the skill requirements will be stated in terms of "traits" (e.g., loyalty, resourcefulness, initiative) and thus lead even more directly to criteria and predictors which are oriented toward underlying predispositions.

RELATIONSHIP TO OTHER ISSUES

The consistency notion has direct relevance for a number of research issues that appear frequently in the selection and prediction literature. One important implication is that selection research should focus on individuals to a much greater extent than it has. That is, there should be more emphasis on intraindividual consistency of behavior. In their insightful discussion of the criterion problem, Ghiselli and Haire (1960) point out that intraindividual criterion performance sometimes varies appreciably over time, that is, is "dynamic." They give two examples of this phenomenon. However, after an exhaustive review of the literature, Ronan and Prien (1966) concluded that a general answer to the question, "Is job performance reliable?" is not really possible with present data. They go on to say that previous research has not adequately considered the relevant dimensions that contribute to job performance and very few studies have actually used the same criterion measure to assess performance at two or more points in time. In the absence of much knowledge concerning the stability of relevant job behaviors it seems a bit dangerous to apply the classic validation model and attempt to generalize from a one-time criterion measure to an appreciable time span of job behavior. Utilizing the consistency notion confronts the problem directly and forces a consideration of what job behaviors are recurring contributors to effective performance (and therefore predictable) and which are not.

In addition, the adoption of signs as predictors in the context of the classic model has undoubtedly been a major factor con-

tributing to the lack of longitudinal research. It makes it far too easy to rely on concurrent studies, and an enormous amount of effort has been expended in that direction. Emphasis on behavior samples and behavior consistency requires that a good deal more attention be devoted to the former, along with very explicit consideration of the crucial parameters of a longitudinal study.

The moderator or subgrouping concept also seems an integral part of the consistency approach. The basic research aim is to find subgroups of people in a particular job family for whom behavior on a particular performance dimension is consistent. Subgrouping may be by individual or situational characteristics but the necessity is clear and inescapable. Only within such subgroups is longitudinal prediction possible.

Lastly, the process the authors are advocating demands a great deal in terms of being able to specify the contextual or situational factors that influence performance. It is extremely important to have some knowledge of the stimulus conditions under which the job behavior is emitted such that a more precise comparison to the predictor behavior sample can be made. Because of present difficulties in specifying the stimulus conditions in an organization (e.g., Sells, 1964), this may be the weakest link in the entire procedure. However, it is also a severe problem for any other prediction scheme, but is usually not made explicit.

It is important to note that the authors' notion of a consistency model does not rest on a simple deterministic philosophy and is not meant to preclude taking account of so-called "emergent" behaviors. Relative to "creativity," for example, the question becomes whether or not the individual has ever exhibited in similar contexts the particular kind of creative behavior under consideration. If a similar context never existed, the research must investigate creative performance and outputs obtained in a test situation which simulates the contextual limitations and requirements in the job situation.

An additional advantage of the consistency approach is that a number of old or persistent problems fortunately appear to dissipate, or at least become significantly diminished. Consider the following:

1. Faking and response sets—Since the emphasis would be on behavior samples and not on self-reports of attitudes, beliefs, and interests, these kinds of response bias would seem to be less of a problem.

2. Discrimination in testing—According to Doppelt and Bennett (1967) two general charges are often leveled at tests as being discriminatory devices:

(*a*) Lack of relevance—It is charged that test items are often not related to the work required on the job for which the applicant is being considered, and that even where relationships can be shown between test scores and job success there is no need to eliminate low-scoring disadvantaged people since they can be taught the necessary skills and knowledge in a training period after hiring.

(*b*) Unfairness of content—It is further maintained that most existing tests, especially verbal measures, emphasize middle-class concepts and information and are, therefore, unfair to those who have not been exposed to middle-class cultural and educational influences. Consequently, the low test scores which are earned are not indicative of the "true" abilities of the disadvantaged. Predictions of job success made from such scores are therefore held to be inaccurate.

The examination of past behaviors similar in nature to desired future behavior, along with their contextual ramifications, plus the added techniques of work samples and simulation devices encompassing desired future behavior, should markedly reduce both the real and imagined severity of problems of unfairness in prediction.

3. Invasion of privacy—The very nature of the consistency approach would seem to almost entirely eliminate this problem. The link between the preemployment or prepromotion behavior and job behavior is direct and obvious for all to see.

CONCLUDING COMMENTS

The preceding discussion is meant to be critical of the concepts of predictive and concurrent validity. Nothing that has been said here should be construed as an attack on construct validity, although Campbell (1960) has pointed out that reliability and validity are also frequently confused within this con-

cept. Neither do the authors mean to give the impression that a full-scale application of the consistency model would be without difficulty. Using available criteria and signs of assumed underlying determinants within the framework of the classic model is certainly easier; however, for long-term gains and the eventual understanding of job performance, focusing on the measurement of *behavior* would almost certainly pay a higher return on investment.

Some time ago, Goodenough (1949) dichotomized this distinction by referring to signs versus samples as indicators of future behavior. Between Hull's (1928) early statement of test validities and Ghiselli's (1966) more recent review, almost all research and development efforts have been directed at signs. Relatively small benefits seem to have resulted. In contrast, some recent research efforts directed at samples seem to hold out more promise. The AT&T studies, which used ratings of behavior in simulated exercises (Bray & Grant, 1966), and the In-basket studies reported by Lopez (1965) are successful examples of employing behavior samples with management and administrative personnel. Frederiksen (1966) has reported considerable data contributing to the construct validity of the In-basket. In addition, Ghiselli (1966) has demonstrated that an interview rating based on discussion of specific aspects of an individual's previous work and educational history had reasonably high validity, even under very unfavorable circumstances. In a nonbusiness setting, Gordon (1967) found that a work sample yielded relatively high validities for predicting final selection into the Peace Corps and seemed to be largely independent of the tests that were also included as predictors.

Hopefully, these first few attempts are the beginning of a whole new technology of behavior sampling and measurement, in both real and simulated situations. If this technology can be realized and the consistencies of various relevant behavior dimensions mapped out, the selection literature can cease being apologetic and the prediction of performance will have begun to be understood.

REFERENCES

BRAY, D. W., & GRANT, D. L. The assessment center in the measurement of potential for business management. *Psychological Monographs*, 1966, **80**(17, Whole No. 625).

CAMPBELL, D. T. Recommendations for APA test standards regarding construct, trait, and discriminant validity. *American Psychologist*, 1960, **15**, 546–553.

CRONBACH, L. J. *Essentials of psychological testing.* (2nd ed.) New York: Harper & Row, 1960.

DOPPELT, J. P., & BENNETT, G. K. Testing job applicants from disadvantaged groups. *Test Service Bulletin* (No. 57). New York: Psychological Corporation, 1967. Pp. 1–5.

DUNNETTE, M. D. A modified model for test validation and research. *Journal of Applied Psychology*, 1963, **47**, 317–323.

DUNNETTE, M. D. *Personnel selection and placement.* Belmont, Calif.: Wadsworth, 1966.

FREDERIKSEN, N. Validation of a simulation technique. *Organizational Behavior and Human Performance*, 1966, **1**, 87–109.

GHISELLI, E. E. *The validity of occupational aptitude tests.* New York: Wiley, 1966.

GHISELLI, E. E., & HAIRE, M. The validation of selection tests in the light of the dynamic character of criteria. *Personnel Psychology*, 1960, **13**, 225–231.

GOODENOUGH, F. *Mental testing: Its history, principles, and applications.* New York: Holt, Rinehart & Winston, 1949.

GORDON, L. V. Clinical, psychometric, and work sample approaches in the prediction of success in Peace Corps training. *Journal of Applied Psychology*, 1967, **51**, 111–119.

GUION, R. M. Synthetic validity in a small company: A demonstration. *Personnel Psychology*, 1965, **18**, 49–65.

HULL, C. L. *Aptitude testing.* New York: Harcourt, Brace & World, 1928.

LOPEZ, F. M., JR. *Evaluating executive decision making: The In-basket technique.* New York: American Management Association, 1965.

ODIORNE, G. S., & MILLER, E. L. Selection by objectives: A new approach to managerial selection. *Management of Personnel Quarterly*, 1966, **5**(3), 2–10.

RONAN, W. W., & PRIEN, E. P. *Toward a criterion theory: A review and analysis of research and opinion.* Greensboro, N. C.: Richardson Foundation, 1966.

SELLS, S. B. Toward a taxonomy of organizations. In W. W. Cooper, H. J. Leavitt, & W. W. Shelly, II (Eds.), *New perspectives in organization research.* New York: Wiley, 1964.

SMITH, P. C., & KENDALL, L. M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 1963, **47**, 149–155.

WALLACE, S. R. Criteria for what? *American Psychologist*, 1965, **20**, 411–417.

WECHSLER, D. *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corporation, 1955.

(Received October 20, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 5, 377–379

EFFECT OF DEFECT RATE ON INSPECTION ACCURACY

DOUGLAS H. HARRIS

Autonetics Division of North American Rockwell Corporation
Anaheim, California

80 inspectors inspected materials containing defects at 4 different rates. The inspection task employed had been shown in a series of previous studies to be representative of a variety of visual inspection tasks. Performance measures were obtained and analyzed to determine the relationship between defect rate and inspection accuracy. Findings indicated that inspection accuracy decreases with reductions in defect rate. The percentage of true defects detected significantly decreased and the percentage of false reports significantly increased.

Although product inspections are widely relied upon in industrial operations to provide information used in the control of product quality, relatively little is known about the conditions that affect inspection accuracy. There are a number of factors in the inspection task which, under certain circumstances, may operate to increase or decrease the accuracy of the resulting quality information. For example, Harris (1966) found that the complexity of the inspection task had a significant negative effect on inspection accuracy. Even when an unlimited amount of inspection time was provided, the percentage of defects detected was significantly negatively correlated with the complexity of the item inspected.

The objective of this study was to determine the effect of defect rate on inspection accuracy. Most production programs have a typical learning curve associated with them. Early in the program product defect rate is high, but quality typically improves, until later on the defect rate is relatively low. Other circumstances, such as variations in production schedules, may also cause the defect rate to vary in a predictable way. Should inspections be conducted in the same manner regardless of the defect rate? This depends, of course, on the relationship between defect rate and inspection accuracy.

Since visual inspection tasks have some elements in common with visual monitoring tasks requiring the detection of infrequent signals appearing at irregular intervals, previous research on vigilance performance may suggest the general nature of this relationship. One would predict from vigilance research that inspection accuracy would decrease with a decrease in defect rate (Buckner & McGrath, 1963). However, there are differences between typical production inspection tasks and the monitoring tasks which have appeared in the studies on vigilance performance. Consequently, there is little in vigilance research results to indicate the specific nature of the relationship between defect rate and accuracy for an inspection task.

METHOD

A representative inspection task was developed and materials prepared which included defects at four different defect rates. Materials at each defect rate were inspected by a sample of naïve *S*s. Inspection accuracy was measured in terms of the percentage of defects detected and the percentage of false reports made.

Inspection Task

A paper-and-pencil inspection task was developed. The task was similar to that employed in a previously developed selection test for visual inspection work (Harris, 1964). In a series of validation studies, performance on this task was found to be significantly correlated with performance on a wide variety of visual inspection tasks. The correlations are listed in Table 1.

TABLE 1

CORRELATIONS BETWEEN PERFORMANCE ON TEST TASK AND VARIOUS INSPECTION TASKS

Inspection task	No. inspectors	Correlation with test task
Electronic chassis	11	.39
Inertial instruments	8	.86*
Module assemblies	19	.58*
Circuit boards	27	.64*
Microelectronic devices	10	.78*
Photomasks	15	.52*

* $p < .05$.

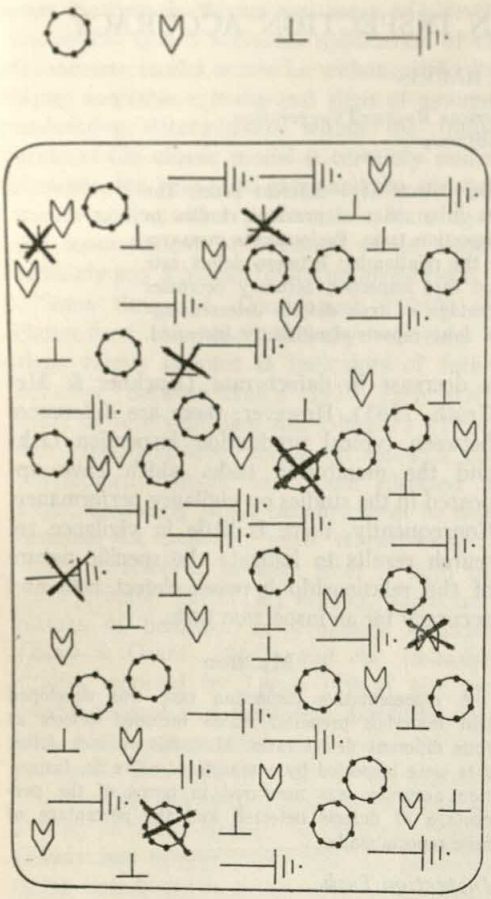


FIG. 1. Sample inspection sheet with defects indicated.

The inspection task consisted of detecting defects in sheets like the one shown in Figure 1. The enclosed area contained 80 items: 20 each of the four types shown in their correct form and orientation at the side of the sheet. Defective items were indicated by an X made over them with a colored pencil.

Four different 40-sheet sets were prepared, one for each of the four different defect rates. The layout of each sheet was identical to the one shown in Figure 1. The starting point in the preparation of materials was drafting a defect-free sheet and printing 160 copies of it. Then a basic set of 32 defective items was drafted and copies printed. Defects were inserted by gluing defective items over appropriate items on the defect-free sheets. From these masters, 20 copies of each 40-sheet set were printed.

Defects were randomly assigned to form four 40-sheet sets with defect rates of 0.25%, 1%, 4%, and 16%. Since each set contained a total of 3,200 items (40 sheets × 80 items per sheet), these percentages correspond to 8, 32, 128, and 516 defects per set of 40 sheets. In preparing the 1%, 4%, and 16% defect rates, the 32 defective items were assigned once, 4 times, and 16 times to the 40-sheet set. For the 0.25% defect rate, four subsets of 40 sheets were each assigned 8 of the basic set of 32 defective items. In this manner, each defective item appeared an equal number of times under each defect-rate condition. A table of random numbers was used to assign defects to sheets.

Procedure

Inspections were made under the four defect-rate conditions by a total of 80 naïve inspectors. Inspectors were randomly assigned to the 4 conditions, 20 per condition. The inspectors were paid student volunteers from California State College at Fullerton and Fullerton Junior College. Each inspector was given samples of the types of defects to be detected and instructed to X over each defect found on each sheet. Inspectors were encouraged to take as much time as necessary to perform accurate inspections. They were given no instructions about the number of defects to expect; they were instructed, however, to ask for clarification on any item about which they were undecided. In this manner, the monitor performed a role similar to an inspection supervisor on the job.

Inspection accuracy was measured in terms of defects detected and false reports made. For each inspection condition, the number of defects detected was divided by the number of defects present to give the percentage of defects detected. The percentage of false reports was computed by dividing the number of defects reported which were not actually defects by the total number of defects reported.

RESULTS AND DISCUSSION

Inspection accuracy decreased with reductions in defect rate. As shown in Figure 2,

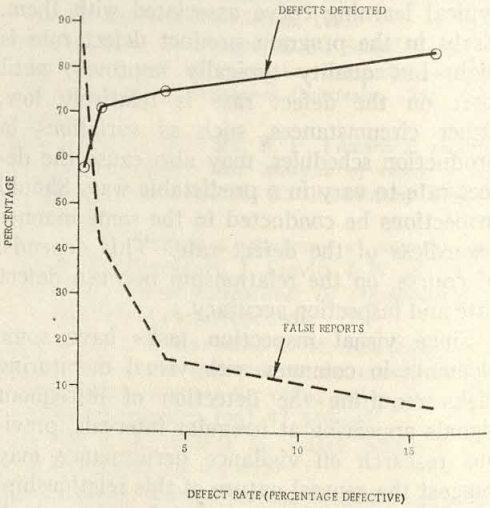


FIG. 2. Relationship between defect rate and inspection accuracy.

the percentage of defects detected decreased slightly between defect rates of 16% and 1%, but dropped sharply between 1% and 0.25%. False reports, the second indicator of inspection accuracy, became more frequent with reductions in defect rate. The percentage of false reports increased at an accelerated rate as the defect rate approached zero. The differences among the four defect rates for the percentages of defects detected and for the percentages of false reports were tested by means of the Kruskal-Wallis one-way analysis of variance by ranks (Siegel, 1956). Differences were statistically significant beyond the .05 level for both the percentages of defects detected and percentages of false reports.

These results demonstrate a potential problem in using information obtained from visual inspections. When a relatively high quality product is being produced (the percentage of defective characteristics is about 1 or less), inspections may not be very accurate. The percentage of defects detected is likely to be low and the percentage of false reports is likely to be high. As a consequence, inferences made about the true quality level from visual inspections may be in error.

These findings have implications for the management of inspection operations. Supervisors should be aware that this tendency exists and, thus, be better prepared to counteract the effect of this factor. For example, as quality level increases, considerations might be given to fewer, more intensive inspections made through the use of sampling plans. Much better information might be obtained at less cost, for example, from inspections of a 10% random sample of the product than from inspections of 100% of the product.

REFERENCES

BUCKNER, D. N., & MCGRATH, J. J. (Ed.) *Vigilance: A symposium*. New York: McGraw-Hill, 1963.

HARRIS, D. H. Development and validation of an aptitude test for inspectors of electronic equipment. *Journal of Industrial Psychology*, 1964, 2, 29–35.

HARRIS, D. H. Effect of equipment complexity on inspection performance. *Journal of Applied Psychology*, 1966, 50, 236–237.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

(Received October 21, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 5, 380–385

PREFERENCE SCALING OF CAREERS AND ORGANIZATIONS[1]

DANIEL N. BRAUNSTEIN AND GEORGE H. HAINES

College of Business Administration, University of Rochester

Paired-comparison scaling procedures were applied to an investigation of preferences for 5 types of careers and 5 types of organizations. Scaling was accomplished for 10 work-satisfaction criteria, using 4 samples, 2 of college students and 2 of college preparatory high school students. The median correlation of preferences between samples was .75; internal consistencies of preferences were higher for college students. The order of preferences for both careers and organizations depended upon the criteria.

The time-worn debate between the "businessman" and the "humanist" regarding the virtues of a business career continues to appear in the popular press. As usual, both assert that their point of view conforms to current student preferences (Blough, 1966; Drucker, 1965). This assertion is possible because accurate data bearing on the question have not been available, and traditional vocational preference theory has not addressed itself directly to the problem.

For today's college student this debate may be more practical than at any time in the past. Employment opportunities in teaching, government, research, and social service have expanded greatly since 1929 (Fuchs, 1965). As the first step in research on this problem, this study attempts to assess student preferences for a limited number of careers, using 10 criteria, and to also assess preferences for the type of organization in which these careers might be practiced.

Most vocational preference theory tries to account for the development of "interests"; for example, perhaps a graduating senior wants to become an engineer because he enjoyed playing by himself with an erector set rather than playing doctor with the little girl next door. This theory often focuses on the biographical antecedents of measured interest patterns; for example, a current area of research studies the development of a "self" concept, an increasing awareness of one's own role (Hunt, 1967). Although vocational interests can be fairly well measured, neither the theories nor empirical research have resulted in a clear statement of the decision process which a student undertakes as graduation nears (Holland, 1964).

This study attempts to measure attitudes which may affect career decisions by determining whether widely held preferences exist for careers and organization types in which careers may be pursued. It is hypothesized that, for the same criteria, preference patterns for business organizations may be considerably different than those for a business administration career.

The issue of preferences raises the question of the basis of career decisions. One model which has been proposed by economists asserts that (assuming people have appropriate information) career decisions are based on an assessment of the present value of future earnings of alternatives at the time the career decision is made (Wilkinson, 1966). The present value model (in its simplest form) states that when people are nearly indifferent in their estimates of nonmonetary rewards, such as a desire for power or leisure, they choose a career on the basis of higher expected earning power. A greater difference in monetary rewards is needed in order for the decision to be swung by the monetary factor, the less the indifference to the nonmonetary factors. Empirical evidence on individual career decisions does not exist to support or deny the idea that people behave in a fashion similar to that proposed above. Data show that aggregate shifts in numbers choosing careers usually follow shifts in salary offers. However, the question as to how individuals actually make their decisions is not answered

[1] This study was supported by a grant from Security Trust Corporation, Rochester, New York. Parts of this paper were presented at the meeting of the Midwestern Psychological Association, Chicago, May 1967.

adequately since the data are purely post hoc for the aggregate. For present purposes it will be assumed that people are able to provide a qualitative ranking which includes responses intrinsic to the work activity and the monetary reward therein.

For example, it could be argued that each person is able to estimate (qualitatively) some just noticeable difference in income. Within the range of this income difference, nonmonetary matters are used to decide between alternative career patterns. Outside this range, trade offs between income and other criteria may be made. For some people it may be that possible choices, after income is taken into account, will be very wide. However, very accurate information on how these trade offs are made is not available.

This study is set up to measure separately the preferences for careers and organizations on a basis of 10 criteria related to job satisfaction.

METHOD

Criteria

Ten criteria for comparing careers and organizations were established after consideration of a number of reviews of the job-satisfaction literature (Bass, 1965; Herzberg, 1966; Vroom, 1964). The criteria are:

(*a*) there are better opportunities to accomplish something worthwhile.
(*b*) there are better advancement possibilities.
(*c*) there will be more situations where you can apply what you learned in school.
(*d*) your success would depend *least* on lucky circumstances.
(*e*) you would earn the highest salary.
(*f*) you would have more important responsibilities.
(*g*) you would be most recognized for your contributions.
(*h*) you would receive supervision from most competent superiors.
(*i*) you would have the most secure future, with steady employment.
(*j*) you would have the best chance of working with people with whom you could build lasting friendships.

The 10 criteria were added to the following pair of questions:

"On your first job in which [type of organization] do you think [Criteria *a* through *j*]."
"On your first job doing which [type of job] do you think [Criteria *a* through *j*]."

Paired-Comparison Instrument

The following is a list of organizations and careers which were compared. The careers are ones which bright students may consider, and usually involve graduate school.

CAREERS

Business Administrator
Lawyer
Research Scientist
Professor
Clinical Psychologist

ORGANIZATIONS

Federal Government
College Faculty
Small Business
Non-Profit Service
Large Corporation

The exact wording of every question asked will be found in Tables 1 and 2. Given these criteria and the above list of careers and future employers, a paper-and-pencil form containing a series of 20 blocks of 10 paired-comparison items was developed. All possible pairs of either careers or organizations were included in a block. Each member of a pair was placed under columns designated "A" or "B," followed by a blank space under a column labeled "choice." The order of pairs within a block was determined by using a table of random numbers. The 10 organizational comparison blocks were followed by the 10 career comparison blocks. Two sample pairs using items not included in the actual comparisons were included in the instructions. The *S*s were instructed to work rapidly. Most completion times, including response to a biographical information form, were between 30 and 40 min. This instrument is at present being used in a number of validation studies.

Subjects

The instrument was administered to four samples of *S*s. Two of the groups were high school seniors ($n = 72$ and 25) who have indicated their intent to go to college. The high schools were located in middle-class suburban communities. The third group consisted of 116 freshmen students at a medium-sized eastern university with a highly selective admissions policy who were enrolled in a freshman psychology course. A fourth sample consisted of 69 upperclassmen at the same university enrolled in an experimental psychology and a statistics course.

RESULTS

Computation of Preference Values

The individual preferences were evaluated according to a procedure suggested by Guilford (1954).[2]

[2] A program to perform these computations written by W. R. Johnson for the IBM 7074 is available from the authors.

TABLE 1.—PREFERENCE

Question	Federal Government				College faculty			
	High school		University		High school		University	
	Sample 1	Sample 2	Fresh-men	Upper-classmen	Sample 1	Sample 2	Fresh-men	Upper-classmen
(a) On your first job, in which do you think there are better opportunities to accomplish something worthwhile?	−.027	.061	−.107	−.210	.302	.229	.512	.647
(b) On your first job in which do you think there are better advancement possibilities?	.169	.156	.119	.190	−.209	−.038	−.152	−.198
(c) On your first job, in which do you think there will be more situations where you can apply what you learned in school?	−.111	.143	−.117	−.237	.253	.451	.753	.732
(d) On your first job, in which do you think your success would depend *least* on lucky circumstances?	.044	−.061	−.095	−.057	.369	.311	.583	.504
(e) On your first job, in which do you think you would earn the highest salary?	.394	.534	.194	.015	−.093	−.062	−.066	−.039
(f) On your first job, in which do you think you would have more important responsibilities?	.107	.209	−.259	−.371	.341	.417	.295	.314
(g) In your first job, in which do you think you would be most recognized for your contributions?	−.387	−.409	−.585	−.568	.162	.090	.392	.397
(h) In your first job, in which do you think you would receive supervision from most competent superiors?	.351	.233	−.034	−.174	.302	.250	.562	.555
(i) In your first job, in which do you think you would have the most secure future, with steady employment?	.406	.679	.645	.583	.168	.192	.098	.167
(j) In your first job, in which do you think you would have the best chance of working with people with whom you could build lasting friendships?	−.803	−.965	−.593	−.630	.577	.424	.721	.887

Note.—Data are *Z*-score values representing the average preference for each item compared to all other items. A zero (neutral) probability of preference, when compared to the other items.
High school: Sample 1 $n = 72$; Sample 2 $n = 25$. University: Freshmen $n = 116$, Upperclassmen $n = 69$.

The average proportion of preferences for every career and organization was calculated. This average proportion was converted to a scale on a standard score basis, with 0 as a neutral point. Positive scale values indicated a higher proportion of preferences than .5; negative, a lower proportion.

Preference Values for Organizations

The data for the organizational preferences are given in Table 1. There are considerable differences between the criteria as to the type of organization most preferred. Federal Government is generally preferred for the security criterion; college faculty for accomplishment, application of school work, and least dependence on luck; large corporation for advancement and salary; non-profit service and small business shared preferred positions for the recognition and social criteria.

Preference Values for Careers

The data for the career preferences are given in Table 2. Business administrator is strongly preferred for its advancement opportunities. Among the high school students, lawyer is viewed as offering high starting salaries, while the college *Ss*' preferences for lawyer are lower on the salary criterion. Research scientist is preferred for recognition and competent supervision and, by all except the college upperclassmen, for accomplishment. The upperclassmen strongly prefer clinical psychologist for the accomplishment and responsibility; it is not highly preferred by the others (the upperclassmen in the present sample were psychology majors). College professor is strongly preferred, surprisingly, on the social criterion.

Internal Consistency

If such an obtained scale of preference values is to be maximally useful, one must assume that the scale results can be used *as if* the so-called "strong axiom of revealed preference" held. That is, if A is preferred to B and B to C, then C ought not to be preferred to A. If it is, this triad would be inconsistent.

VALUES FOR ORGANIZATIONS

Large corporation				Non-profit service organization				Small business			
High school		University		High school		University		High school		University	
Sample 1	Sample 2	Fresh-men	Upper-classmen	Sample 1	Sample 2	Fresh-men	Upper-classmen	Sample 1	Sample 2	Fresh-men	Upper-classmen
−.176	.203	.139	−.279	.108	−.066	−.008	.427	−.205	−.020	−.258	−.628
.431	.451	.488	.452	−.590	−.696	−.502	−.325	.183	.127	.048	−.153
.000	.069	−.068	−.201	−.250	−.728	−.256	.229	.107	.065	−.311	−.524
.000	.019	−.229	−.188	−.268	−.167	.038	.228	−.147	−.102	−.296	−.488
.885	.972	1.154	1.199	−1.183	−1.481	−1.279	−1.118	−.005	.037	−.002	−.014
−.476	−.750	−.455	−.529	−.137	−.148	.131	.306	.164	.272	.287	.280
−.379	−.402	−.467	−.376	.305	.327	.324	.349	.285	.395	.336	.198
.299	.150	.281	.299	−.612	−.493	−.421	−.252	−.338	−.144	−.391	−.387
.254	.341	.223	.165	−.514	−.996	−.535	−.218	−.315	−.121	−.432	−.688
−.646	−.352	−.752	−.629	.561	.418	.424	.373	.311	.476	.200	.001

value indicates that the item is preferred half the time when compared to the others. The higher the score, the higher the

The presence or absence of this condition (consistency of behavior) was tested using the Dobell (1965) method.

Consistency allows construction of a meaningful scale. If this condition, which might be called an internal consistency condition, is not met, then it is not possible to infer back from the scale the actual choices which would be observed in paired-comparison choice situations. It should be noted that a lack of internal consistency does not necessarily imply unreliability—preference choices may be reliably inconsistent.

Internal consistency was greater for the college students than the high school students. Eliminating the counting of ties as "inconsistent," all organizational criteria and all but two career criteria (competent supervision and security) were consistent for college students.

Consistency between Samples

Product-moment correlations between samples were computed for all pairs of the four samples of data.

Tables 3 and 4 give the sample intercorrelations. They range from .52 to .84; the mode and median is .75. The range for careers is greater than that for organizations. All are significant beyond the .01 level.

DISCUSSION

This article began with the concept that preferences both for careers and organizations must be measured concurrently because pref-

TABLE 3

CORRELATION MATRIX: ORGANIZATIONS

Sample	1	2	3	4
1. High school, Sample 2[a]	1.00	0.61	0.59	0.78
2. University upper-classmen[b]	0.61	1.00	0.75	0.81
3. University freshmen[c]	0.59	0.75	1.00	0.76
4. High school, Sample 1[d]	0.78	0.81	0.76	1.00

[a] $N = 25$.
[b] $N = 69$.
[c] $N = 116$.
[d] $N = 72$.

TABLE 2.—PREFERENCE

Question	Business administrator				Lawyer			
	High school		University		High school		University	
	Sample 1	Sample 2	Fresh-men	Upper-classmen	Sample 1	Sample 2	Fresh-men	Upper-classmen
(a) On your first job, doing which do you think there are better opportunities to accomplish something worthwhile?	−.800	−.817	−.869	−.877	−.162	−.069	−.072	−.349
(b) On your first job, doing which do you think there are better advancement possibilities?	.605	.399	.646	.506	.048	.056	.428	.086
(c) On your first job, doing which do you think there will be more situations where you can apply what you learned in school?	.035	−.060	−.492	−1.096	−.172	−.238	−.198	−.498
(d) On your first job, doing which do you think your success would depend *least* on lucky circumstances?	−.185	−.162	−.100	−.179	.000	.141	−.162	−.308
(e) On your first job, doing which do you think you would earn the highest salary?	−.137	−.138	.560	.390	.647	.802	.476	.309
(f) On your first job, doing which do you think you would have more important responsibilities?	−.489	−.432	−.333	−.381	.129	.001	.107	−.215
(g) In your first job, doing which do you think you would be most recognized for your contributions?	−.508	−.373	−.469	−.563	.130	−.008	.117	−.022
(h) In your first job, doing which do you think you would receive supervision from most competent superiors?	−.085	−.061	−.183	−.231	−.085	−.230	−.076	−.130
(i) In your first job, doing which do you think you would have the most secure future, with steady employment?	.098	.187	.159	.098	.037	.144	−.022	−.134
(j) In your first job, doing which do you think you would have the best chance of working with people with whom you could build lasting friendships?	−.101	.026	−.337	−.530	−.204	−.063	.046	−.224

Note.—Data are *Z*-score values representing the average preference for each item compared to all other items. A zero (neutral) probability of preference, when compared to the other items.
High school: Sample 1 $n = 72$; Sample 2 $n = 25$. University: Freshmen $n = 116$; Upperclassmen $n = 69$.

erence patterns for business organizations may be considerably different than for business administration careers. The results tend to support the hypothesis. However, the differences which exist for any criterion would, the authors assert, be of small importance, if in fact the preference patterns of students for careers and organizations were internally inconsistent. Vocational interest theory, combined with studies of the value of future income, must consider these social biases in order to clarify the career-decision-making process.

TABLE 4

CORRELATION MATRIX: CAREERS

Sample	1	2	3	4
1. High school, Sample 2[a]	1.00	0.52	0.59	0.84
2. University upper-classmen[b]	0.52	1.00	0.75	0.63
3. University freshmen[c]	0.59	0.74	1.00	0.75
4. High school, Sample 1[d]	0.84	0.63	0.75	1.00

[a] $N = 25$.
[b] $N = 69$.
[c] $N = 116$.
[d] $N = 72$.

This study also investigated the possibility that there would be substantial agreement among students regarding preferences which may have an effect upon career decisions. The four samples, two of high school students and two of college students, yielded preferences which showed a high amount of agreement with each other. This agreement indicates that the preferences may have a considerable amount of stability across other such samples.

The most important implication of this study is that neither business careers nor business organizations are universally downgraded by students in comparison to the other alternatives available to them, although general business administration careers may have

VALUES FOR CAREERS

Research scientist				Clinical psychologist				College professor			
High school		University		High school		University		High school		University	
Sample 1	Sample 2	Fresh-men	Upper-classmen	Sample 1	Sample 2	Fresh-men	Upper-classmen	Sample 1	Sample 2	Fresh-men	Upper-classmen
.609	.563	.438	.418	.196	.271	.205	.760	.162	.052	.297	.236
.097	.049	−.334	−.016	−.358	−.402	−.430	−.219	−.392	−.102	−.310	−.358
.121	.213	.207	.649	−.319	−.316	.056	.675	.334	.403	.426	.384
−.013	.165	−.112	−.023	.006	−.082	−.007	.211	.192	−.063	.382	.300
.154	.025	−.102	.157	−.084	.123	−.284	−.298	−.580	−.791	−.620	−.557
.031	.199	−.004	−.100	.137	.171	.120	.720	.191	.061	.153	−.025
.633	.575	.592	.439	−.126	−.029	−.147	.154	−.128	−.165	−.093	−.006
.369	.269	.398	.646	−.238	−.040	−.098	−.132	.040	.062	−.040	−.153
−.289	.005	−.323	−.066	−.115	−.273	.074	−.079	.268	−.063	.271	.179
−.285	−.395	−.237	−.280	−.014	−.346	.010	.448	.605	.778	.518	.586

value indicates that the item is preferred half the time when compared to the others. The higher the score, the higher the

difficulty in attracting these students compared to the professions for most career objectives. The existence of commonly held stereotypes, such as found in this research, may indicate to businessmen, teachers, and guidance counselors that our culture may be transmitting some strong influences to students as they make their career decisions.

REFERENCES

BASS, B. M. *Organizational psychology*. Boston: Allyn and Bacon, 1965.

BLOUGH, R. M. Business can satisfy the young intellectual. *Harvard Business Review*, 1966, **44**(1), 49–57.

DOBELL, A. R. A comment on AYC Koo's "An empirical test of revealed preference theory." *Econometrica*, 1965, **33**(2), 451–455.

DRUCKER, P. F. Is business letting young people down? *Harvard Business Review*, 1965, **43**(6), 49–55.

FUCHS, V. R. The growing importance of the service industries. *Journal of Business*, 1965, **38**, 344–373.

GUILFORD, J. P. *Psychometric methods*. (2nd ed.) New York: McGraw-Hill, 1954.

HERZBERG, F. *Work and the nature of man*. Cleveland: World, 1966.

HOLLAND, J. C. Major programs of research on vocational behavior. In H. Borow (Ed.), *Man and the world of work*. Boston: Houghton Mifflin, 1964.

HUNT, R. A. Self and other semantic concepts in relation to choice of a vocation. *Journal of Applied Psychology*, 1967, **51**, 242–246.

VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.

WILKINSON, B. W. Present values of lifetime earnings for different occupations. *Journal of Political Economy*, 1966, **74**, 556–572.

(Received October 24, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 5, 386–393

TASK COHERENCE, TRAINING TIME, AND RETENTION INTERVAL EFFECTS ON SKILL RETENTION[1]

JAMES C. NAYLOR,[2] GEORGE E. BRIGGS, AND WALTER G. REED[3]

Ohio State University

Amount of training, secondary task coherence, and length of retention interval, each at 2 levels, were evaluated in terms of long-term skill retention effects. The criterion task was composed of a 3-dimensional tracking task (primary task) and a 9-event monitoring task (secondary task). Retention loss varied inversely with amount of training and with secondary task coherence (the latter only under the lesser amount of training) for both tasks. Absolute retention levels varied directly with the training and task coherence variables and inversely with retention interval. From these and previous data, task coherence emerges as an important variable in skill acquisition and retention.

In their review of the literature on skill retention, Naylor and Briggs (1961) concluded "that the retention of an arbitrary sequence of responses (whether verbal or motor) is less than the retention of responses which occur in meaningful sequences or with patterned organizations [p. 6]." At first sight the results of some recent tracking research by Trumbo, Noble, Cross, and Ulrich (1965) and by Noble, Trumbo, Ulrich, and Cross (1966) seem to be opposite to the above conclusion: They found greater *losses* in performance over the retention interval for more predictable tasks. However, the absolute level of performance at retention test was better for the more coherent tasks, thus supporting the above conclusion. In a recent summary and review of their research, Noble and Trumbo (1967) suggest that the greater retention losses for the more coherent tasks may be due primarily to a deterioration in timing accuracy over the retention interval. This interpretation is consistent with a study by Naylor and Briggs (1963) which found rehearsal on the temporal aspects of a task much more effective than rehearsal on the spatial aspects in terms of retention performance. Therefore, while deterioration occurs in both the temporal and the spatial aspects of learned responses, the former is the more crucial loss to skill.

In a more recent study, Trumbo, Noble, and Swink (1967) utilized a tracking situation as a primary task and added a secondary task (verbal anticipation or guessing a sequence of numbers) during retention testing (Experiment I). Coherence was varied systematically in both tasks, and, while primary task performance during retention was inferior when the subject (*S*) experienced the two tasks simultaneously, there was no *differential* effect of the level of predictability of the secondary task on primary task performance. As in their previous work, the authors found differential retention performance as a function of primary task coherence.

It seems odd that secondary task coherence would not produce differential effects on primary task performance: From the concept of a central, limited channel for information processing (Broadbent, 1958), one would predict a differential influence of both primary and secondary task coherence provided, of course, that the experimenter (*E*) selected sufficiently different levels of secondary task predictability. In other words, it is conceivable that Trumbo et al. (1965) did not differentially load the channel in their two secondary task conditions where the obtained average uncertainty in the stimulus sequence was .542 and 1.882 bits/signal.

The present research examined the influence

[1] This research was carried out in the Human Performance Center and was supported by the Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, under Contract No. AF 33(616)-7269 with the Ohio State University Research Foundation. Further reproduction is authorized to satisfy the needs of the United States Government.

[2] Now at Purdue University.

[3] Now at Armstrong Cork Company, Lancaster, Pennsylvania.

of two levels of secondary task coherence on primary (tracking) task performance during both learning and retention. Both tasks were more complex than those utilized by Trumbo et al. (1967), and an attempt was made to assure a significant separation of secondary task coherence levels. However, both of the coherence levels used here provided *S* with the opportunity for errorless performance on the secondary task, in contrast to the Trumbo et al. (1965) research where *S* could improve with practice on the secondary task but could not achieve errorless performance except by chance.

It was predicted that both acquisition and retention performance would vary directly with secondary task coherence for both the primary and the secondary tasks. In addition to secondary task coherence, the amount of training (2 or 3 wk.) and the retention interval (1 or 4 wk.) were varied systematically to determine if the above prediction would hold across these two highly pertinent variables in retention research.

METHOD

Apparatus

The *S* performed two tasks simultaneously: a three-dimensional tracking task and a nine-event monitoring task. The primary (tracking) task presented *S* with three center-reading voltmeter displays which provided system error information, and he attempted to null out such error by manipulating a single control column via left-right, fore-aft, and rotation responses. Each of the tracking dimensions involved rate dynamics and the control-display gain was .025 ips of display pointer movement per degree of control device movement. The control was spring loaded in all three dimensions at approximately 1 oz/deg of arc, and a single sinusoid of .6 cpm served as the tracking input signal to all three dimensions of the primary task.

The secondary (monitoring) task presented *S* with nine rows of stimulus and response devices to indicate nine events. Within each row there were (from left to right) three response buttons, plus an amber- and a red-jewel light. An event occurred every 6 sec. If an amber light occurred, *S* was to press the button labeled "OK" for that event; if a red light appeared, he first pressed the "Emergency" button (which extinguished the red light and activated the amber light), and then he was required to press the "OK" button; and if no light occurred at the expected time, *S* pressed first a button labeled "Check" and this activated the red light, which in turn required him to press first the "Emergency" and then the "OK" buttons. In all cases a correct sequence of responses within 4 sec. of event onset resulted in an amber light being locked in; an incorrect sequence resulted in a red light being locked in.

The secondary task was located to the left of the tracking display and the separation was such as to require *S* to look away from the tracking display in order to respond, although the secondary display lights were of sufficient intensity to be clearly apparent in peripheral vision.

Secondary task coherence was manipulated in terms of the spatial sequence of events: In the highly predictable conditions the spatial sequence was 1, 2, 3, 4, 5, 6, 7, 8, 9, where 1 refers to the top pair of lights, 2 to the next pair, and so on; and for the low coherence condition the spatial sequence was 1, 5, 2, 9, 8, 3, 6, 7, 4. Stated in terms of the amount of change within the spatial sequence of events, the high coherence sequence involved 0 bits of uncertainty while the low predictability sequence involved 2.16 bits of uncertainty per change.

Design and Procedure

Table 1 summarizes the treatments for each of the eight experimental conditions. The latter resulted from factorial combinations of (*a*) training time (2 or 3 wk.), (*b*) retention interval (1 or 4 wk.), and (*c*) monitoring task coherence (high and low). Each group was filled with 16 male undergraduates who volunteered for service and were paid $1.00 per session. Group assignment was made on the basis of total tracking task performance during the sixth session.

All *S*s experienced part-task training through the fifth session with whole-task training and retention testing thereafter. Except for the first session, which involved four 70-sec. trials on each of the two tasks, there were twelve 70-sec. trials per session. The tracking task was always practiced as a whole (three-dimensional) task, and there was no change in the primary task from the first to the last session. However, the secondary task was increased in complexity across the first five sessions in terms of the number of amber, red, and no-light events: during the first two sessions the stimulus events were always red or

TABLE 1

EXPERIMENTAL CONDITIONS FOR THE VARIOUS RETENTION GROUPS

Group	Amount of training (in wk.)	Task organization	Retention interval (in wk.)
1	2	High	1
2	2	High	4
3	2	Low	1
4	2	Low	4
5	3	High	1
6	3	High	4
7	3	Low	1
8	3	Low	4

another; during the third session *S* experienced first one no-light event (Trials 2 and 3) and then two no-light events (Trials 5 and 6); and in the fourth session this was increased first to three no-light events (Trials 2 and 3) and then to four no-light events (Trials 5 and 6). Beginning with the second trial of the fifth session, there were four no-light events on every trial including those during the retention test. The number of out- and another-light events were scheduled so as to be equal across the several sessions.

Tracking performance was scored over the final 60 sec. of each 70-sec. trial which coincided with the time during which the monitoring task events occurred (the latter being an event every 6 sec. starting 10 sec. after the beginning of a trial). Integrated absolute tracking error was recorded for each dimension and summed for the three dimensions. These data were transformed from their original voltage units to units of linear extent (inches) on the tracking display scale (Briggs, 1964).

Monitoring task performance was scored in three ways: (*a*) total response time for the nine events on each trial, (*b*) number of omissive errors (failure to respond at the proper time or to complete the required response sequence within the first 4 sec. of event onset), and (*c*) commissive errors (responses in excess of the required number and/or incorrect responses for the particular event).

The instructions to *S* stressed the need to maintain minimal tracking error and to respond quickly and accurately to the monitoring task. Each *S* was shown his scores at the end of each session to emphasize these goals.

RESULTS

Tracking Performance

Figures 1 and 2 provide the average tracking performance levels during both training and retention test for the 2- and the 3-wk. training conditions, respectively. The data for

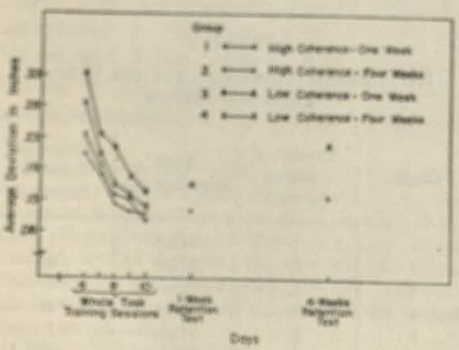


FIG. 1. Tracking performance during training and retention testing of the four groups receiving 2 wk. of training.

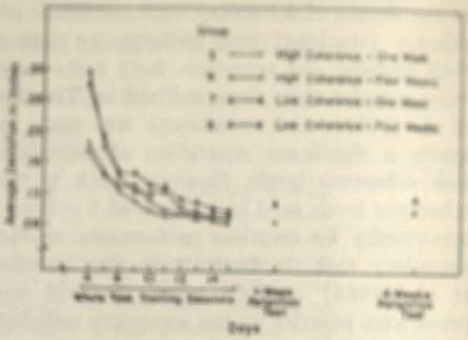


FIG. 2. Tracking performance during training and retention testing of the four groups receiving 3 wk. of training.

the first five sessions are not included, as a part-task training procedure was in effect during that time and the monitoring task was being increased in complexity. Each data point in Figures 1 and 2 represents tracking performance averaged over 12 trials and 16 *Ss*, a total behavioral sample of 192 min. per point.

Training. It is apparent that secondary task coherence affected primary task performance as predicted: Groups 1 and 2 were superior to Groups 3 and 4 (Figure 1) while Groups 5 and 6 were superior to Groups 7 and 8 (Figure 2), on the average, during training. Further, it is apparent by comparing Figures 1 and 2 that terminal performance was superior for the 3-wk. training groups. These observations are supported by the results of an analysis of variance which was applied to the average performance of each *S* on his last block of four training trials: both amount of training ($F = 5.92$, $df = 1/124$) and secondary task coherence ($F = 5.25$, $df = 1/124$) were statistically significant at $p < .05$. However, the interaction of these two variables was not significant ($F < 1.0$).

Retention. From the above, groups differed at the end of training with particularly superior tracking performance by Groups 5 and 6 which experienced 3 wk. of training under coherent secondary task conditions. Therefore, the retention data were analyzed in two ways: (*a*) Difference scores were calculated for each *S* (last training session minus retention test performance) which correct for the initial

TABLE 2

ANALYSES OF VARIANCE OF DIFFERENCE SCORES (TRAINING-RETENTION) FOR ALL PERFORMANCE MEASURES

Source	*df*	Tracking		Omissive errors		Commissive errors		Response time	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Amount of training (T)	1	12399.21	22.57**	2.0125	11.21**	2.1321	1.98	38.8962	3.29
Task organization (O)	1	1660.34	3.02	.2993	1.67	.6296	—	.0132	—
Retention interval (I)	1	101.55	—	.0037	—	13.1072	12.18**	.6877	—
T × O	1	3012.80	5.48*	1.2940	7.21**	1.5269	1.42	14.0183	1.18
T × I	1	1664.63	3.03	.0000	—	.2794	—	1.5706	—
O × I	1	188.16	—	.0001	—	.5279	—	27.6308	2.33
T × O × I	1	1797.02	3.27	.0469	—	.7503	—	1.5065	—
Residual error	120	549.41		.1795		1.0764		11.8582	

* $p < .05$.
** $p < .01$.

group differences and permit, therefore, an examination of the effects of the experimental variables on retention performance independently of their effects on training performance; and (*b*) the uncorrected retention performance scores were analyzed to permit an examination of the effects of the experimental variables on the absolute (rather than relative) levels of retention performance.

These two ways of analyzing retention performance permit one to examine two different aspects of the influence of the independent variables on skill retention. In the case of relative scores one is concerned with the *loss* or *amount of deterioration* in performance over a time interval, while with the absolute scores one is interested in what *remains* after a time interval.

The results of these analyses are shown in Tables 2 and 3. In terms of relative loss (Table 2) those groups which experienced only 2 wk. of training deteriorated more during the retention periods than did the groups which received 3 wk. of training, and, while secondary task coherence did not exert an overall significant effect on the amount of loss, the significant interaction of Coherence × Training indicates that a greater loss occurred for those *Ss* who received only 2 wk. of training on the less predictable secondary task (Groups 3 and 4) than for any other *Ss*, including those who also received only 2 wk. of training but did so on the more coherent version of the secondary task (Groups 1 and 2). It is interesting to note that retention interval (1 versus 4 wk.) did not produce dif-

TABLE 3

ANALYSES OF VARIANCE OF ABSOLUTE RETENTION SCORES FOR ALL PERFORMANCE MEASURES

Source	*df*	Tracking		Omissive errors		Commissive errors		Response time	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Amount of training (T)	1	9080.47	18.45**	1.9404	16.10**	5.2407	4.09*	36.7440	4.22*
Task organization (O)	1	6419.86	13.05**	1.6608	13.78**	.0069	—	11.2694	1.30
Retention interval (I)	1	2263.62	4.62*	.1891	1.57	8.8200	6.89**	1.6562	—
T × O	1	1546.37	3.16	1.1101	9.21**	.9940	—	10.3285	1.19
T × I	1	380.53	—	.0657	—	.0443	—	.0935	—
O × I	1	168.13	—	.0741	—	.8224	—	27.2137	3.13
T × O × I	1	421.59	—	.0587	—	.4705	—	19.5625	2.25
Residual error	120	492.10		.1205		1.2801		8.7001	

* $p < .05$.
** $p < .01$.

TABLE 4

MEAN PERFORMANCE PER SUBJECT FOR EACH OF THE EIGHT EXPERIMENTAL GROUPS DURING THE LAST BLOCK OF TRAINING TRIALS AND AT RETENTION TEST FOR EACH OF THE THREE MONITORING TASK METRICS

Group	Omissive errors		Commissive errors		Response time (in sec.)	
	Training	Retention	Training	Retention	Training	Retention
1	.18	.14	1.02	.92	7.61	6.91
2	.25	.17	.71	1.36	8.24	8.89
3	.27	.46	.79	1.07	8.74	9.77
4	.30	.68	.60	1.59	7.87	8.35
5	.08	.08	.67	.77	7.98	7.24
6	.05	.11	.69	1.05	6.74	7.55
7	.08	.12	.58	.34	7.04	7.41
8	.13	.16	.60	1.17	8.41	7.44

ferential losses either as a main effect or in interaction with either of the other two independent variables.

In terms of absolute retention performance on the tracking task, Table 3 shows that all three independent variables were statistically significant as main effects. Thus, not only was superior retention performance obtained under high secondary task coherence and after 3 wk. of training (compared to that under low coherence and after only 2 wk of training), respectively, but also performance after a 1-wk. retention interval was superior to that after 4 wk.

In summary, then, amount of training influenced both the amount of deterioration in tracking performance and the absolute performance level at retention test; secondary task coherence differentially influenced the amount of deterioration only for groups which received 2 wk. of original training, although there was an overall effect in terms of absolute performance levels at retention test; and, while there was no greater loss in performance over a 1-wk. retention period than over 4 wk., the final average performance level after 1 wk. was superior to that after the longer retention interval.

Monitoring Task Performance

Table 4 provides summaries of the performance levels of the eight groups during training (final block of four trials) and retention test. It can be seen that very few errors of either type occurred and, in general, *S* was more apt to commit errors (extra responses) than to fail to respond. Tables 2 and 3 provide the results of analyses of variance applied to the monitoring task scores during retention test.

Omissive errors. During the final training trials there were significantly more omissive errors by *S*s who experienced only 2 wk. of training (Groups 1–4) than by those who experienced 3 wk. of training ($F = 5.81$, $df = 1/124$); however, unlike the case with tracking performance data, secondary task coherence did not differentially affect omissive errors during training. In terms of the amount of deterioration over the retention interval, the omissive error data revealed the same pattern as with the tracking data: greater losses occurred in the 2-wk. training groups than in the 3-wk. groups, and the Training × Coherence interaction was statistically significant (see Table 2). From Table 3 it can be seen that omissive errors followed the same pattern as found for tracking performance, except that retention interval was not significant and the Training × Coherence interaction was significant at $p < .05$. Again, the pattern of this interaction was the same as that found for the relative loss scores on the tracking and monitoring tasks (Table 2): the worst performance was achieved under the 2-wk. training, low coherence conditions (Groups 3 and 4) with fairly comparable performance for all other groups.

Comissive errors and response time. These

two measures of monitoring task performance were less sensitive to the effects of the independent variables than either the omissive error or the tracking error scores. For example, there were no significant differences with either set of scores during training and there was only one significant F ratio from the two retention analyses of the response time scores: amount of training for the absolute retention data (see Table 3). Those few cases of significant F ratios for both commissive error and response time data do reveal results entirely consistent with those of the tracking error and omissive error data; for example, better retention performance occurred after 3 than after 2 wk. of training in terms of both commissive errors and response time (Table 3), and there were more commissive errors after 4 wk. than after 1 wk. of retention (Table 3). It is interesting to note that the commissive error data were the only ones to show a significant effect of retention interval for *both* the relative loss and the absolute performance analyses (see Tables 2 and 3).

This lack of sensitivity of the response time and commissive error scores to the independent variables may well have been due to the delay in feedback which S experienced with regard to these two aspects of performance: the S obtained such information only as supplementary feedback at the end of each daily session when E provided a debriefing. The feedback on tracking error was immediately and continuously available from the three tracking displays, and a failure to respond appropriately to each event in the monitoring task was signaled almost immediately by a red light on the monitoring display.

DISCUSSION

The above results indicate that each of the independent variables influenced both original acquisition and retention tracking performance in a straightforward and predictable manner, with two exceptions. Thus, during original acquisition, tracking performance was superior after 3 wk. compared to that after 2 wk. of training, and superior tracking performance occurred when accomplished with simultaneous practice on the more coherent than with the less predictable secondary task. During retention, tracking performance, in terms of absolute scores (Table 3), was significantly better (*a*) after 3 rather than 2 wk. of original training, (*b*) when accomplished simultaneously with the more coherent secondary task, and (*c*) after 1 wk. rather than after 4 wk. of retention. Finally, in terms of performance loss scores (Table 2), there was less loss following 3 wk. of original training, and there was less tracking performance loss when S experienced the more predictable secondary task, *but* this effect held only for those Ss who received only 2 wk. of original training. The latter was one of the two exceptions noted above: secondary task coherence did not appear as a main effect but as an interactive effect with amount of training in terms of retention loss.

The other exception was that retention interval did not differentially affect performance *loss* although it did influence absolute retention performance, as noted above. This latter exception is actually a bit puzzling, since the magnitude of differences between means due to the retention interval variable was actually *larger* in the difference score analysis than it was in the absolute score analysis. One possible explanation for obtaining nonsignificance due to retention interval in the difference score analysis is that the within-cell variance was markedly increased in this analysis due to the unreliability of individual S's performance. Since the variance of a set of difference scores is $s_D^2 = s_A^2 + s_B^2 - 2r_{AB}s_As_B$, a low correlation between the training and retention test scores of individuals within a group would result in a rather large s_D^2, which in turn would result in a corresponding increase in the MS error term used in the analysis of variance.

Essentially the same pattern of results was found with performance on the secondary (monitoring) task. In terms of omissive errors, the only discrepancy between results on the two tasks was that retention interval did not influence performance as a main effect for either the retention loss data or the absolute retention scores. Retention interval was a statistically significant effect in terms of commissive errors, however.

It follows, then, that the major prediction was supported, on the whole, by the results of this experiment: Both during training and at retention test, superior primary task performance was obtained when it occurred simultaneously with the more coherent version of the secondary task, and performance on the secondary task itself during retention (but not during training) was superior under the more coherent than under the less predictable version. The latter result is consistent with those of Trumbo et al. (1965) and Noble et al. (1966), while the results noted here for the retention of tracking performance extend those of Trumbo et al. (1967, Experiment I) by showing that secondary task coherence can affect primary task performance in a predictable manner.

The most consistently influential and the most powerful independent variable in this experiment was the amount of original training: on both tasks during training and at retention, performance levels were superior for groups which experienced 3 rather than 2 wk. of original training, and the *loss* of proficiency on both tasks was less for the more highly trained groups. Further, the interactions noted in Table 2 of Training × Secondary Task Coherence indicate that not only can one reduce retention losses on both primary and secondary tasks with training, but also the detrimental effect of low coherence can be ameliorated with sufficient original learning.

Surprisingly, the effect of retention interval was less consistent than that of either of the other two independent variables. The fact that retention interval did not interact significantly with secondary task coherence indicates that the effect of the latter is uniform across at least a 4-wk. interval. Logically, one would expect such an interaction to be significant, and visual inspection of Figure 1 seems to support this expectation: there appears to be a greater difference between Groups 2 and 4 than between Groups 1 and 3 at retention test. However, again, this observation was not supported by statistical test, and obviously from Figure 2 there is no greater difference between Groups 6 and 8 than between Groups 5 and 7. Had a 6- or an 8-wk. retention interval been utilized instead of the 4-wk. interval, it is quite possible that the expected interaction would have been found. Indeed, one may predict that the interaction of Amount of Training × Secondary Task Coherence × Retention Interval would be significant provided one used *either* lesser amounts of original training *or* greater retention intervals than those used herein.

While there is danger in speculation, the above does provide further (albeit indirect) evidence for the very powerful effects of the amount of training variable: Overlearning can (and apparently did) mask or ameliorate the effects of less powerful independent variables, especially their interactive effects.

The authors' interpretation of the coherence variable, as indicated in the introduction, is that by manipulating task predictability one loads differentially the central information-processing channel. The less coherent a secondary task input, the more the channel must be concerned with the selection and execution of responses for the secondary task, and, therefore, the less the capacity which is available for primary task performance. It follows, then, that during training primary task performance should vary directly with the coherence of the secondary task, and this was the result obtained.

During retention test one encounters not only the above effect of secondary task coherence but also less accurate retention of the input sequence for the less predictable secondary task which compounds the load on the central channel. Presumably, this reduces further the capacity of the channel to deal with the primary task.

The finding that amount of training can ameliorate the effects of secondary task coherence on retention performance (at least in terms of retention *loss*) is consistent with the interpretation of a hierarchical organization of skill (see Chapter 2 of Fitts & Posner, 1967). As (secondary task) skill becomes more automated through practice, it places less load on the central channel, and thus there should be less potential interference to the processing of information on another (primary) task performed simultaneously.

Whatever the theoretical interpretation, task coherence emerges from these and previ-

ous data as an important determinant of skilled performance under long-term retention conditions.

REFERENCES

BRIGGS, G. E. Analog computers. In J. B. Sidowski (Ed.), *Experimental methods and instrumentation in psychology*. New York: McGraw-Hill, 1966.

BROADBENT, D. E. *Perception and communication*. London: Pergamon Press, 1958.

FITTS, P. M., & POSNER, M. I. *Human performance*. Belmont, Calif.: Brooks-Cole, 1967.

NAYLOR, J. C., & BRIGGS, G. E. Long-term retention of learned skills: A review of the literature. USAF ASD Technical Report, No. 61-390, August 1961.

NAYLOR, J. C., & BRIGGS, G. E. Effect of rehearsal of temporal and spatial aspects on the long-term retention of a procedural skill. *Journal of Applied Psychology*, 1963, **47**, 120–126.

NOBLE, M., & TRUMBO, D. The organization of skilled response. *Organizational Behavior and Human Performance*, 1967, **2**, 1–25.

NOBLE, M., TRUMBO, D., ULRICH, L., & CROSS, K. Task predictability and the development of tracking skill under extended practice. *Journal of Experimental Psychology*, 1966, **72**, 85–94.

TRUMBO, D., NOBLE, M., CROSS, K., & ULRICH, L. Task predictability in the organization, acquisition, and retention of tracking skill. *Journal of Experimental Psychology*, 1965, **70**, 252–263.

TRUMBO, D., NOBLE, M., & SWINK, J. Secondary task interference in the performance of tracking tasks. *Journal of Experimental Psychology*, 1967, **73**, 232–240.

(Received October 27, 1967)

Manuscripts Accepted for Publication in the *Journal of Applied Psychology*

Managers' Attitudes toward Communication Episodes: Edward E. Lawler, III,* Lyman W. Porter, and Allen Tennenbaum: Department of Administrative Sciences, Yale University, 2 Hillhouse Avenue, New Haven, Connecticut 06520.

A Correlational-Causal Analysis of the Relationship between Expectancy Attitudes and Job Performance: Edward E. Lawler, III*: Department of Administrative Sciences, Yale University, 2 Hillhouse Avenue, New Haven, Connecticut 06520.

Consistency and Generalizability of Intraindividual Variability: Ralph F. Berdie*: Student Life Studies, Office of the Dean of Students, University of Minnesota, 2001 Riverside Avenue, Minneapolis, Minnesota 55455.

Complex Vigilance: Relevant and Irrelevant Signals: Irwin L. Goldstein,* William A. Johnston, and William C. Howell: Department of Psychology, University of Maryland, College Park, Maryland 20740.

Prediction of Job Success for Hospital Aides and Orderlies from MMPI Scores and Personal History Data: James N. McClelland* and Fen Rhodes: Department of Psychology, California State College, 6101 East Seventh Street, Long Beach, California 90804.

Multivariate Relationships between a Measure of Vocational Interests and a Measure of Vocational Needs: Robert M. Thorndike,* David J. Weiss, Rene V. Dawis: Work Adjustment Project, Industrial Relations Center, University of Minnesota, Minneapolis, Minnesota 55455.

Off-Quadrant Comment: Norman M. Abrahams*: United States Naval Personnel Research Activity, San Diego, California 92152.

Moderation of a Moderator Technique: Quinn McNemar*: Department of Psychology, University of Texas, Mezes Hall 211, Austin, Texas 78712.

Multidimensional Analysis of Worker-Oriented and Job-Oriented Verbs: John Charles Allen*: Department of Psychology, Morrill Hall, University of Maryland, College Park, Maryland 20742.

Ratee Relevance in Peer Nominations: Frank T. Passini* and Warren T. Norman: Department of Psychology, The University of Michigan, Ann Arbor, Michigan 48104.

Experience and Prior Probability in a Complex Decision Task: Michael H. Strub*: The Human Performance Center, Ohio State University, 404-B West 17th Avenue, Columbus, Ohio 43210.

Marginal Productivity Procedure for Staff Selection: Purnell H. Benson*: 123 Milligan Place, South Orange, New Jersey 07079.

---

* Asterisk indicates author for whom the address is supplied.

(Continued on page 422)

Journal of Applied Psychology
1968, Vol. 52, No. 5, 394–398

USE OF IMPORTANCE AS A WEIGHTING COMPONENT OF JOB SATISFACTION

PATRICIA SMITH MIKES AND CHARLES L. HULIN [1]

University of Illinois, Urbana

Hypotheses dealing with the rated importance of various aspects of the job were tested on data obtained from 660 male and female workers in Montreal, Quebec. Importance used multiplicatively to weight satisfaction scores did not increase the predictability of turnover during an 11-mo. period over that obtained by using satisfaction measures alone. No interaction between satisfaction and importance predicting turnover was obtained. The general conclusion was that importance has little value in a prediction situation involving job attitudes and behavior.

Strong and consistent relationships between employees' responses to attitude questionnaires and job behavior have only rarely been obtained. It has been implied frequently that those aspects of the job which are perceived as more important by the worker have a greater influence on the behavioral and overall affective responses of the worker to his job than do less important aspects. However, most job-attitude questionnaires fail to ask employees about aspects regarded as more or less important.

Several investigators (Ewen, 1967; Friedlander, 1965; Glennon, Owens, Smith, & Albright, 1960; Weiss, Dawis, England, & Lofquist, 1964a, 1964b) have used importance measures in conjunction with satisfaction in an effort to demonstrate theoretically and empirically more appealing measures. All but one concluded that a consideration of importance is necessary. Ewen, however, concluded that importance added little. His results showed that a summed composite score weighted differentially for importance ratings was no more strongly related to a measure of overall satisfaction than was an unweighted composite of the same satisfaction measures.

The research reported in this paper presents what the authors believe to be more definitive evidence on the importance of importance. If importance is a vital component in the assessment of job satisfaction then overall job satisfaction is dependent on the worker's attitude toward the various aspects of the job, and also on the perceived importance of these aspects to him. His general reaction to his job would depend more heavily on the feeling he holds toward those aspects which he considers as more important than on his feelings toward those aspects which are less important. A weighted composite measure should yield a measure of overall satisfaction with the job which is more highly related to behaviors such as turnover or absences than an unweighted composite measure. The model proposed is a multiplicative interaction of importance and satisfaction with a linear combination of the products.

The *S*'s composite job attitude as measured by the product of his importance rating and satisfaction score for each job aspect summed over all aspects will correlate more highly with termination decisions than will either the summed importance ratings or an unweighted sum of the satisfaction measures.

Evidence supporting this hypothesis would indicate that the perceived importance of job aspects is necessary if respectable predictive ability is to be achieved. Negative evidence would indicate that Ewen's (1967) basic conclusions can be generalized to the prediction of behavioral responses. (Such negative evidence could mean that the importance of any aspect of the job is reflected in the reported satisfaction with that aspect of the job. That is, only those job aspects which are important to *S* will be responded to by him with a satisfaction score which deviates substantially from the neutral point. Unimportant job

[1] These data were gathered in conjunction with and reported as part of the senior author's master's thesis submitted to the University of Illinois. The authors would like to thank George Graen and Harry Triandis for the comments on an earlier draft of the manuscript.

aspects would receive essentially neutral satisfaction scores and would enter the predictive equations with very small effective weights.)

Negative evidence could also be interpreted as a reflection of the reliability of a composite measure as compared to the reliability of a simple sum. Every time one measures a variable some error is introduced. If satisfaction with each job characteristic is measured with a reliability of .80 then one would expect the sum of the five importance scores to have reliability considerably in excess of .80.[2] However, this reliability would be greater than the reliability of the same five satisfaction scores if they were multiplied by a fallible measure of importance before they were summed. Even if the measures of importance used have high (but less than perfect) reliability, the resulting sum of Importance × Satisfaction would have an overall reliability much below the reliability of the simple sum. The maximum obtainable correlation between the simple sum and any criterion would be greater than the maximum obtainable correlation between a composite measure and the criterion. If all the correlations were corrected for attenuation it is possible that these corrected correlations would indicate the *theoretical* superiority of the composite sum of job satisfaction. However, the problem cannot be removed by demonstrating the theoretical superiority of one measure over another. It is an unfortunate fact of life that every time one introduces the notion of importance in the study of job satisfaction such importance variables will have to be measured. Importance measures cannot be discarded on the basis of the available data. However, the claims that importance *must* be considered in the study of job satisfaction are unfounded since such claims cannot ignore the problems of reliability.

METHOD

Subjects

The *S*s for this study were the corporate office personnel of a large international corporation with its corporate offices located in Montreal, Quebec. In July 1966 all of the employees were requested to take part in a job-satisfaction survey which the company was conducting in cooperation with the second author. The questionnaires were given to the employees by departmental representatives after the representatives had met with the investigator and had the questionnaire explained to them. The employees were asked to complete them at home or, if they thought they would have 30 uninterrupted minutes, to do them at work. All employees were asked to participate. Completed questionnaires were placed in envelopes, sealed by the respondents, and returned either to the investigator or sent through the mails directly to the University of Illinois. Anonymity was guaranteed to all respondents and they were assured that their individual responses would not be made known to the company. Names were not requested on the questionnaires but for many of the employees their job title and department (which were requested) would be enough to identify the respondent. Of the 800 employees who worked in the home offices and who were present during the period of the study, 670 (83%) returned usable questionnaires.

Variables

Satisfaction with five aspects of the job (work done, pay level, promotional opportunities, supervision, and co-workers) was assessed by means of the Job Descriptive Index (JDI) (Smith, 1967) which is a cumulative-point adjective checklist. In addition, the General Motors Faces Scale (Kunin, 1955) was used to assess job satisfaction with the same five job aspects. In this particular study a 6-point version of the original 11-point scale was used. The six faces used consisted of two scowling faces, one neutral face, and three smiling faces. A measure of overall job satisfaction was also obtained. The same 6-point version of the General Motors Faces Scale was used to assess overall satisfaction.

Importance was measured by a scale designed to give interval scale values of importance. A description of this scale may be found in Hulin and Smith (1967) or Mikes (1967). This scale provided ratings of importance of six aspects of the job (interest and difficulty of work, pay, promotion, supervision, co-workers, and working conditions). Satisfaction with work was used in conjunction with both the rated importance of interest and difficulty of work and of working conditions since the satisfaction measures may not adequately discriminate between these aspects of the job.

Background variables (job level, sex, age, tenure) were obtained from the face sheet of the questionnaire which asked for this and other information. Job level was quantified from job listings of the company by matching the job title reported on the face sheet with the job evaluation scale provided by the personnel department of the company involved. All of these measures were obtained in July 1966.

The reasons given by the terminators for quitting were recorded during the initial stages of this study. The plan was to eliminate those *S*s from the sample

[2] The use of one of the Kuder-Richardson formulas would be inappropriate to estimate the reliability of the sum since the measures of satisfaction with the separate job components are not homogeneous parts of a longer test.

who were "unavoidable" quits (pregnancy, husband leaving town, etc.). However, the distinction between avoidable and unavoidable quickly became meaningless. Within the first month two girls who were pregnant reported to the personnel department. One terminated. Her questionnaire was placed in the unavoidable category. The second girl did not quit but only requested a maternity leave and planned to return to work. It was impossible to justify the exclusion of the first girl. The decision was made to include all terminators regardless of stated reason for quitting.

RESULTS

By May 1967, 6 males and 16 females who had participated in the study and who could be identified had quit their jobs. No correlational analyses were performed on the male sample since the number of cases was too small. Analyses for the female sample were performed on a group of 48 *S*s, 16 terminators and 32 nonterminators who were matched on age, length of service, and job level with the terminators. Biserial correlations were computed between the sum of JDI scores and turnover (scored 1 for termination, 0 for nontermination), the sum of the Faces scores with job aspects and turnover, the sum of importance-weighted JDI scores and turnover, the sum of importance-weighted Faces scores and turnover, the measure of overall job satisfaction and turnover, and the sum of importance scores and turnover. The results of this analysis are presented in Table 1.

The hypothesis was not confirmed. The unweighted sums of the satisfaction scores alone proved to be as highly related to termination decisions as the sums of the importance-weighted satisfaction scores. The least accurate predictor was the sum of importance scores alone.

TABLE 1

BISERIAL CORRELATION BETWEEN PREDICTOR VARIABLES AND TURNOVER

Variable	Correlation with turnover
Sum of JDI scores	−.46
Sum of JDI scores weighted by importance	−.33
Sum of Faces scores	−.41
Sum of Faces scores weighted by importance	−.39
Job-in-general Faces score	−.34
Sum of importance scores	−.22

Note.—Females, $N = 48$.

One possible explanation for the lower predictive validity of the importance-weighted satisfaction scores was that the multiplicative model used did not distinguish between job aspects high in importance but low in satisfaction and job aspects low in importance but high in satisfaction. The first case would be an indicator of future turnover but the latter case would not. To test this problem of lack of symmetry, the female terminators and their matched control *S*s were dichotomized at the medians of satisfaction with work itself and the importance of work itself. The same procedure was followed using the 6 males and their 12 matched control *S*s. In each case the medians were computed separately for the male and female *S*s and therefore different median values were used for each of the two samples. Satisfaction with work itself was the variable selected for this analysis because it had the highest correlation with turnover of any of the five JDI measures. The dichotomized male and female turnovers and their matched controls were then combined preserving the median splits which had been obtained from the separate analyses of the two samples. This dichotomization and combining of male and female workers yielded a four-cell table based on a total N of 66 (22 terminators and 44 matched nonterminators). The ratio of terminators to total cell numbers in each of the four cells was then computed. The interactive model of importance and satisfaction would be supported if the 2×2 table indicated a high turnover ratio in the high-importance–low-satisfaction cell, a low ratio in the high-satisfaction–high-importance cell, and intermediate turnover ratios in the other two cells. These data are presented in Table 2. No statistical analyses were done on these data but an inspection indicates the existence of main effects due to satisfaction and importance but no interaction effects are present.

DISCUSSION

The multiplicative model presented in this paper received little empirical support. The sum of satisfaction scores with unit weights did as well as the composite sum of satisfac-

tion scores when weighted by importance in the prediction of turnover. The intuitively appealing idea that importance is important in the study of job satisfaction received no support. The most immediate conclusion which could be reached on the basis of these data would be that there is no need to use importance and an interactive model to link satisfaction with the different aspects of the job to the behavioral responses to the job since the simpler model assuming unit weights is more parsimonious and is at least as valid.

There are alternative conclusions which could be reached on the basis of these data. It is possible that the interactive model of job satisfaction is valid but that the measures of satisfaction used in this study are invalid. Such a conclusion lacks force, however, when the validity data for the JDI and Faces measures of satisfaction are examined. These data (Smith, 1967) indicate high convergent and discriminant validity for the JDI. The numerous studies which have appeared recently using the General Motors Faces Scale of overall satisfaction have yielded results which would be impossible if the Faces Scale lacked validity (Ewen, Smith, Hulin, & Locke, 1966; Graen, 1966; Hulin & Smith, 1967). It is also possible that the multiplicative model of job satisfaction is basically valid but the measure of importance of the various job aspects employed lacks validity. The data presented by Mikes (1967) indicate high scale reliability for this measure. The data pertaining to the convergent and discriminant validity are less impressive. Nevertheless, these data indicate at least an adequate amount of convergent validity.

One psychometric consideration must be discussed. What was done in this study was to weight a limited number of variables using idiosyncratic, individually determined weights. Using such idiosyncratic weights, the resulting sum of the weighted variables will have a correlation in excess of .90 with the unweighted sum of the same variables (Ewen, 1967). This extremely high correlation, of course, reduces drastically the chances of demonstrating the superiority of one measure or another.

However, it should be stressed that this psychometric consideration is not unique to this particular study. Any study which uses

TABLE 2

ANALYSIS OF TURNOVER RATES AS PREDICTED BY AN INTERACTION OF IMPORTANCE AND SATISFACTION

Satisfaction with work itself	Importance of work itself		Total
	High	Low	
High	5/23 (22%)	3/10 (30%)	8/33
Low	3/10 (30%)	11/23 (49%)	14/33
Total	8/33	14/33	22/66

Note.—Numerators in cells are number who quit and denominators indicate total cell frequency.

importance as a weighting measure to weight measures of job satisfaction will have the same problem associated with it. This means that claims for the superiority of weighted composite measures of job satisfaction are very likely unjustified.

It is also possible that had this study used a different behavioral measure as its criterion the results would have been more promising for a multiplicative model. Termination decisions were used as the dependent variable because this is the one behavioral response which has indicated consistent and high relationships with affective responses to the job (Hulin, 1966, 1968; Weitz & Nuckols, 1953). The use of productive behavior as a dependent variable would have created many problems since the previous studies on job satisfaction which attempted to link affective responses to productive behavior have not indicated anything like a consistent or significant relationship (Brayfield & Crockett, 1955; Vroom, 1964).

In addition, more promising results might have been obtained had other prediction models been used to combine the satisfaction and importance data. The use of other ad hoc models (such as a multiple-regression prediction equation) would clearly have taken advantage of the negative relationships between importance ratings and turnover, the negative relationships between satisfaction and turnover, and the zero association between satisfaction and importance to enhance the predic-

tion of turnover. However, with only 48 *S*s in the present analysis sample such analyses could have proved misleading.

It should also be pointed out that this study does not stand alone. The research of Ewen (1967) while using a different dependent variable (overall job satisfaction) yielded results no more promising for an interactive model of job satisfaction. It would appear that, every time the intuitively appealing use of importance in combination with satisfaction has been investigated in order to determine the *validity* of such a model, negative results have been obtained.

In summary, it appears that the use of importance measures as multiplicative weights in a model of job satisfaction does not in fact yield the relationships between a composite measure of job satisfaction and behavior, or between a composite measure of job satisfaction and overall satisfaction which are stronger, more consistent, or more theoretically appealing. The simpler model of job satisfaction assuming unit weights for the measures does as well.

REFERENCES

BRAYFIELD, A. H., & CROCKETT, W. H. Employee attitudes and employee performance. *Psychological Bulletin,* 1955, **52**, 396–424.

EWEN, R. B. Weighting components of job satisfaction. *Journal of Applied Psychology,* 1967, **51**, 68–73.

EWEN, R. B., SMITH, P. C., HULIN, C. L., & LOCKE, E. A. An empirical test of the Herzberg two-factor theory. *Journal of Applied Psychology,* 1966, **50**, 437–441.

FRIEDLANDER, F. Relations between the importance and the satisfaction of various environmental factors. *Journal of Applied Psychology,* 1965, **49**, 160–164.

GLENNON, J. R., OWENS, W. A., SMITH, W. J., & ALBRIGHT, L. E. New dimensions in measuring morale. *Harvard Business Review,* 1960, **38**, 106–107.

GRAEN, G. B. Addendum to "An empirical test of the Herzberg two-factor theory." *Journal of Applied Psychology,* 1966, **50**, 442–445.

HOTELLING, H. The generalization of the student's ratio. *Annals of Mathematical Statistics,* 1931, **2**, 360–378.

HULIN, C. L. Job satisfaction and turnover in a female clerical population. *Journal of Applied Psychology,* 1966, **50**, 280–285.

HULIN, C. L. Effects of changes in job-satisfaction levels on employee turnover. *Journal of Applied Psychology,* 1968, **52**, 122–126.

HULIN, C. L., & SMITH, P. A. An empirical investigation of two implications of the two-factor theory of job satisfaction. *Journal of Applied Psychology,* 1967, **51**, 396–402.

KUNIN, T. The construction of a new type of attitude measure. *Personnel Psychology,* 1955, **8**, 65–78.

MIKES, P. S. An investigation of an interactive model of job satisfaction. Unpublished master's thesis, University of Illinois, Urbana, 1967.

SMITH, P. C. The development of a method of measuring job satisfaction: The Cornell Studies. In E. A. Fleishman (Ed.), *Studies in personnel and industrial psychology.* (Rev. ed.) Homewood, Ill.: Dorsey Press, 1967.

VROOM, V. *Work and motivation.* New York: Wiley, 1964.

WEISS, D. J., DAWIS, R. V., ENGLAND, G. W., & LOFQUIST, L. H. Construct validation studies of the Minnesota Importance Questionnaire. *Minnesota Studies of Vocational Rehabilitation,* 16, 1964, Bulletin 41. (a)

WEISS, D. J., DAWIS, R. V., ENGLAND, G. W., & LOFQUIST, L. H. The measurement of vocational needs. *Minnesota Studies of Vocational Rehabilitation,* 16, 1964, Bulletin 39. (b)

WEITZ, J., & NUCKOLS, R. C. The validity of direct and indirect questions in measuring job satisfaction. *Personnel Psychology,* 1953, **6**, 487–494.

(Received November 1, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 5, 399–402

EVALUATING MORTGAGE LOAN RISK

WILLIAM D. BUEL

Byron Harless, Schaffer, Reid & Associates, Inc., Tampa, Florida

Mortgage loan application blanks for matched samples of delinquent and nondelinquent mortgages were statistically analyzed for purposes of identifying items capable of discriminating between the 2 samples. 8 bits of information (4 application blank items and 4 ratios) differentiated delinquent from nondelinquent accounts and, taken in total score form, gave rise to a biserial correlation of .64 ($p < .001$).

Recent efforts by psychologists have demonstrated the validity of weighted credit application blanks. McGrath (1960) reports a maximum biserial correlation of .58 (several weighting methods were compared) between a total score based on 24 weighted credit application blank items and a payment-repossession criterion relative to automobile purchase, while Hassler, Myers, and Seldin (1963) discovered nine items capable of producing a biserial correlation of .71 when taken in total score form relative to differentiating "open" from "special handling" department store accounts. Elsewhere, Myers (1963) describes point-biserial correlations ranging from .26 to .61 for various weighting methods relative to differentiating paid from delinquent accounts and accepted from rejected applicants, respectively.

Literature searches by the present author have disclosed no such efforts relative to mortgage loans. In fact, the "lore" of the mortgage-lending industry suggests that risk evaluation would be relatively difficult because payment on one's home looms more important than payment on less significant items, and therefore elicits a different set of payment motives, motives alleged to cause mortgage installment payments where "other" periodic remittances might be neglected. While such motives may operate, the present study exhibits validity comparable to the validity of credit scoring systems designed for other purposes.

Criterion

So that payment records would be available for parallel samples of delinquent and nondelinquent mortgage loan recipients, this study employed historic data, data arising from the mortgage application blanks of individuals to whom mortgage loans had been extended for purposes of home construction or purchase—judgments about delinquency–nondelinquency could be made only on the basis of real experience.

Examination of the organization's delinquent files suggested an artificially dichotomous criterion.[1] Delinquency was therefore defined as 1 or more month's failure to pay, whereas nondelinquency suggested no payment failures. While such stringent definitions make differences difficult to discover, they were forced upon the investigator by sample-size constraints, that is, no usable data could be spared.

Method

The mortgage loan application blanks to be studied were drawn from files, by date of loan issuance, for the period February 1960 to April 1966. Once 150 mortgage loan application blanks representing delinquent accounts had been selected, 150 representative of nondelinquents were chosen by searching the files immediately following each of the delinquent accounts, selecting only those blanks which could be matched with a delinquent account blank on date (within 30 days), purpose of application (construction or purchase), and present address (town). Accordingly, these matching variables served as their own control and are, across both samples, not necessary of separate statistical treatment.

After selection of 150 delinquent and 150 nondelinquent accounts, every third pair of files (50 delinquent and 50 nondelinquent) was removed so as to provide for a 100-account cross-validation group which paralleled the 200-account item-analysis sample (100 delinquent and 100 nondelinquent accounts) to be used in the analysis of the information on the mortgage loan application blank. Also, with reference to the delinquent group, the item-analysis and cross-validation samples were matched on number of months of delinquency.

[1] Appreciation is extended to Warren E. Backert and James E. Dirst of the Home Building and Loan Association of Aurora, Illinois for permission to publish this study.

Thirty-one quantifiable or categorizable pieces of information were available from each mortgage loan application blank, as follows: amount of monthly payment, amount of loan, address, possession of telephone, availability of pledge, length of mortgage in months, purpose of mortgage, purchase price, down payment, total value of property (home and lot combined), value of lot, value of building, husband's occupation, husband's employer, husband's number of months of service, husband's salary per month, marital status, wife's employer, wife's months of service, wife's salary per month, other income (husband and wife), total salary per month (husband and wife), total income per month (husband and wife), age of husband, age of wife, number of children, total dependents other than self (including wife whether or not she was employed), value of other monthly payments, balance owed on those monthly payments, number of other accounts, and a 4-point credit rating from the local credit bureau.[2] These 31 bits of information were analyzed, via tetrachoric coefficients of correlation,[3] to determine their power of differentiation between the delinquent and nondelinquent accounts in the item-analysis sample.

In addition to the analysis of the items of information described above, 21 of them, either because of their quantitative nature or by logical dictate, lent themselves to inclusion in ratios with other items. Accordingly, a 21×21 paired-comparison matrix was prepared to facilitate consideration of all possible combinations. While 210 ratios could have been developed $\frac{[N(N-1)]}{2}$, logic suggested the elimination of certain of them. For example, the ratio "number of children" to "total dependents" is almost meaningless since it breaks down into a series of constants for mortgagees with like numbers of children and dependents. Further, the intercorrelation of these two values is .9904, indicating overlap and nonindependence. Such examples are not infrequent, and judgment dictated that only 79 ratios be calculated and tested for their significance of differentiation.

RESULTS

Four items of mortgage loan application blank information and four ratios differentiated delinquent from nondelinquent accounts at the .05 level of significance or better ($N = 200$). Those eight items and their tetrachoric correlation coefficients follow: wife employed (+.23), number of children (−.27), number of other accounts (−.28), credit rating (+.45), monthly mortgage payments/age of husband (−.30), amount of loan/total income (−.24), age of husband/number of dependents other than self (+.30), and age of husband/number of other accounts (+.25).

Each of these items and ratios was assigned a set of differential weights ranging from 10 to 0 within each item, weights which while strictly intended for use with qualitative data were chosen for the sake of homogeneity of method across both qualitative and quantitative items (Wherry, 1944). Employing the eight weighted items as a scoring key for use with 100 mortgage loan application blanks held apart from the item analysis and correlating the total scores so derived with the delinquency–nondelinquency dichotomy gave rise to a biserial correlation of .64, significant at the .001 level of confidence. Figure 1 presents the data from which this correlation arises.

Further evidence of validity is connoted by the fact that the 50-person nondelinquent holdout group achieved a mean score of 53.24, whereas the 50-person delinquent holdout group achieved a mean score of only 37.84, a difference of 15.40. Especially in light of the total score standard deviation (15.06 with $N = 100$), the implications of such a mean score difference may be seen. Also, with regard to Figure 1, it is interesting to note that at several points in the distribution 40% more nondelinquents than delinquents equaled or exceeded a particular mortgage loan application blank score. The most dramatic evidence of this sort relates to a score of 50, a point at or above which 64% of the nondelinquent group scored—only 24% of the delinquent group achieved this level.

DISCUSSION

While the personnel evaluator is most usually concerned with positive discrimination (identifying individuals with good performance potential), the mortgage loan evaluator is most interested in discriminating at the

[2] Some items of information (husband's occupation and husband's employer are examples) were analyzed in such fashion as to reflect qualitative rather than quantitative differences, that is, to reflect differences in occupation or employer relative to goodness or poorness of payment behavior. Unfortunately, such categories were numerous and frequencies within them became meaninglessly small and useless.

[3] Tetrachoric coefficients of correlation were employed in all item and ratio analyses because both item and criterion data had been artificially dichotomized by comparing the frequencies of response for a particular interval of item data (the one with the largest delinquent-nondelinquent response difference) against all other intervals of item data.

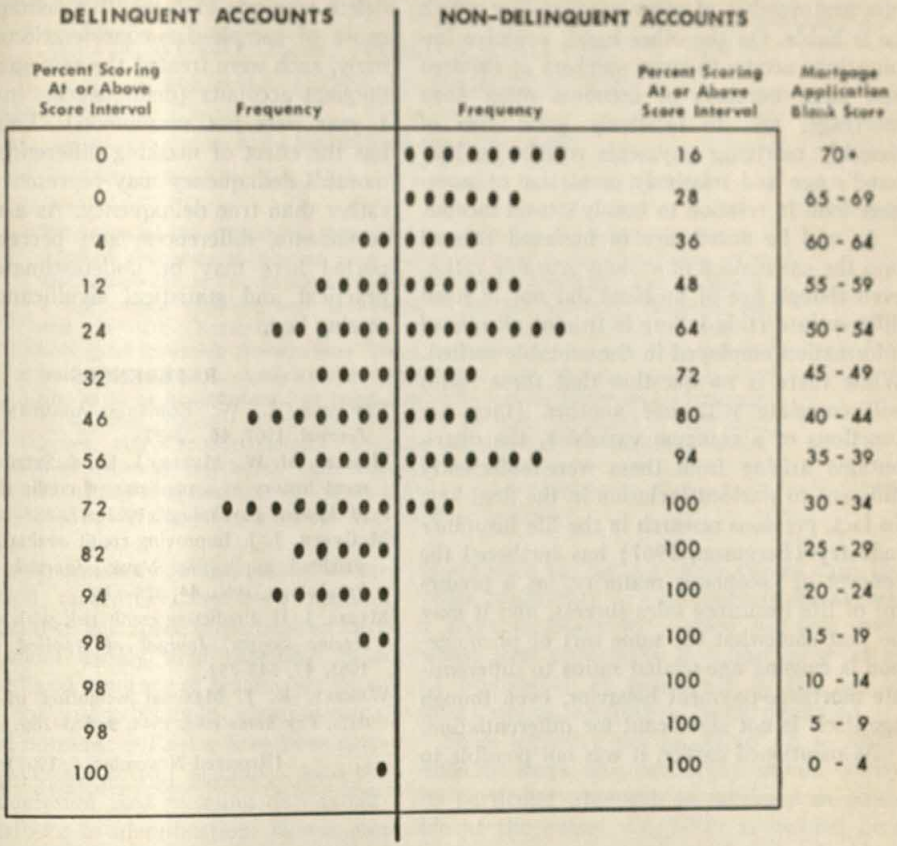


FIG. 1. Results of applying mortgage application blank scoring key to holdout group ($N = 100$).

low end of the scale (identifying the individual with poor payment potential). Accepting this inverted emphasis, an examination of Figure 1 suggests that rejecting individuals achieving mortgage loan application blank scores of 34 or below would result in eliminating individuals most of whom would be poor risks. Lowering screening standards just one score interval (rejecting individuals scoring 29 or below) has even more profound effects in terms of the percentage of known delinquents who would have been rejected.

As is obvious, delinquency and nondelinquency do not comprise opposing cells in a dichotomy. Rather, they are the qualitative extremes of a continuum which can be conceived of in quantitative terms. Perhaps the truth of the matter is that the items discussed here are differentiating with regard to "lateness" of payment (quality), while the criterion defines lateness in terms of number of months of delinquency (quantity). Suffice it to say that at the upper extreme of the lateness continuum, lateness evolves into risk for the loan-granting institution—risk expressed as the cost associated with efforts at collection of back payments, indirect expenses related to the clerical handling costs of delinquent accounts, legal action, and so forth.

The typical mortgage-lending institution is interested in the mortgagee's "success" in maintaining a periodic payment schedule, and the items found valid in this study may be viewed as contributors to or inhibitors of such success. Considering the signs of the item correlations cited earlier, payment success would seem enhanced by a working wife, a good credit rating (a history of payment success), and relatively great age of husband in relation to numbers of individuals dependent upon

him and number of other accounts for which he is liable. On the other hand, negative implications accrue to large numbers of children and large numbers of accounts other than mortgage, and to relatively great size of monthly mortgage payments relative to husband's age and relatively great size of mortgage loan in relation to family's total income.

As will be noted, age of husband entered into the calculation of various scorable ratios, even though age of husband did not in itself differentiate (this failure is true of all bits of information employed in the scorable ratios). While there is no question that these ratios will correlate with one another (they are functions of a common variable), the distributions arising from them were sufficiently different to warrant inclusion in the final key. In fact, previous research in the life insurance industry (Ferguson, 1967) has furthered the concept of "economic maturity" as a predictor of life insurance sales success, and it may be that somewhat the same sort of phenomenon is causing age-related ratios to differentiate mortgage-payment behavior, even though age itself is not significant for differentiation.

As mentioned earlier, it was not possible to delete accounts of 1 month's delinquency because of sample-size considerations. Accordingly, such were treated the same as truly delinquent accounts (delinquencies in excess of 1 year were not uncommon). This practice has the effect of masking differences, since 1 month's delinquency may represent oversight rather than true delinquency. As a result, the coefficients, differences, and percentages reported here may be underestimates of the practical and statistical significance of the scoring key.

REFERENCES

FERGUSON, L. W. Economic maturity. *Personnel Journal*, 1967, **46**, 22–27.

HASSLER, H. W., MYERS, J. H., & SELDIN, M. Payment history as a predictor of credit risk. *Journal of Applied Psychology*, 1963, **47**, 383–385.

MCGRATH, J. J. Improving credit evaluations with a weighted application blank. *Journal of Applied Psychology*, 1960, **44**, 325–328.

MYERS, J. H. Predicting credit risk with a numerical scoring system. *Journal of Applied Psychology*, 1963, **47**, 348–352.

WHERRY, R. J. Maximal weighting of qualitative data. *Psychometrika*, 1944, **9**, 263–266.

(Received November 8, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 5, 403–409

WORD LEGIBILITY AS A FUNCTION OF LETTER LEGIBILITY, WITH WORD SIZE, WORD FAMILIARITY, AND RESOLUTION AS PARAMETERS

R. L. ERDMANN AND A. S. NEAL

International Business Machines Corporation, Los Gatos, California

To determine the relationship between letter legibility and word legibility, individual characters in context were compared with isolated characters. The words included in the stimulus material varied in familiarity and in length. 2 independent variables were introduced to degrade legibility: character size and resolution. Characters of 4 heights were reproduced at each of 20 vertical-horizontal resolution combinations. Word legibility was found to increase with resolution and with character size. Word length, over the range investigated, proved to be unimportant, but word familiarity was an important determinant of legibility. Familiar words were found to be about as legible as individual letters, with a product-moment correlation coefficient of .98. Unfamiliar words were generally less legible than individual letters, with a coefficient of .94. If the degree of familiarity of words is known, word legibility can be predicted with reasonable accuracy on the basis of letter legibility.

In efforts to measure legibility, many different criterion measures have been applied. Viewing distance and speed of perception have been the most widely used threshold measures; speed and accuracy of reading are often the basis for suprathreshold measurements. Words and nonsense syllables have been compared in investigations of legibility, with the general conclusion that meaning and familiarity contribute to identification. Howes and Solomon (1951), studying the influence of familiarity on speed of recognition, found the visual duration threshold (measured tachistoscopically) to be a roughly linear function of word frequency based on the Thorndike-Lorge word count. The product-moment correlations between log word frequency and visual exposure duration ranged from −.76 to −.83.

In subsequent experiments by Solomon and Postman (1952), pronounceable nonsense words were introduced in order to control the "frequency of prior usage." When the relationship between tachistoscopic recognition threshold and frequency was measured, the recognition thresholds were found to vary inversely with word familiarity (as it had been built into the experiment). Howes (1954) interpreted the Solomon-Postman results in terms of response emission probability rather than exposure duration threshold. In his view, word frequency in a large general word count is best interpreted as the average probability of emission by the *S* (which Howes calls the "base probability"). Consequently, he argues, a short exposure will suffice for any word whose base probability is almost as great as its probability of occurrence in the set of words whose legibility is being measured.

Since legibility can be defined in so many different ways, any new study should clarify its particular approach as precisely as possible at the outset. Legibility is defined here as the average percentage of correctly identified letters or words. Legibility is recognized as depending on variables in both the stimulus and the observer. Typically, stimulus material may vary in respect to character size, type style, height-width ratio of characters, stroke width, color, paper surface, and contrast. Variables which include the user are, for example, viewing angle and distance, visual exposure duration, illumination, and visual acuity. There are additional influences, not so readily defined and measured, which are implicit in any attempt to measure legibility. For instance, the *S* is assumed to be familiar with the alphabet from which the test characters are chosen; otherwise the idea of recognition or identification would be meaningless.

Recently, Saltz (1967) has called attention to fundamental problems in the way word frequency—and consequently, word familiarity—has been measured. He points out that

the Thorndike-Lorge word count disregards word meaning in counting frequency, making it impossible to distinguish the two potential influences on legibility. A single spelling ("bear," for example) may represent two meanings, yet its occurrences are counted as if it were a single word. Familiarity, frequency, and meaning overlap in ways which blur the significance of the numbers used to represent each factor.

Whatever the proper interpretation may be, the cumulative evidence indicates that readers recognize frequent or familiar words more quickly than words rare to them, and meaningful words more quickly than nonsense syllables. When distance, rather than speed of perception, is the criterion measure, this continues to hold true. Taylor (1934) found meaningful words to be readable at a greater distance (approximately 50% greater) than nonsense words, whether the printing was black or white. Tinker (1963), investigating differences between uppercase and lowercase letters, found less than a 6% difference in distance, but the difference favored meaningful words for both cases. It might be noted that Tinker's nonsense words were his meaningful words spelled backward; for example, "evoba" comes from "above."

The redundancy in individual letters contributes some degree of familiarity to any combination of letters, whether it is a meaningful word or a nonsense syllable. Tinker (1963) has shown that the upper halves of words are more legibile than the lower halves, and the work of Crooks (1959) gives some indication of the amount of degradation individual forms (other than words) can suffer and still be recognizable. In a sense, the present study attempted to determine how much character degradation can occur before legibility is affected; the main objective, however, was to quantify the relationship between individual letter legibility and word legibility.

In planning the experiment reported here, it was assumed that lowercase letters are more legible in the context of intelligible words than separately, character by character. This assumption is in keeping with work reported by Tinker (1963) which measured legibility by such criteria as fixation frequency, perception time, and perception distance. Information already available (from Erdmann & Neal, 1966) had determined the resolution needed to make individual characters legible (as required in parts lists, for example, and other arbitrary groupings). That resolution was assumed to establish the maximum resolution requirement for all conditions. Therefore, if systematic trends were found between word legibility and resolution, the corresponding relationship between word and letter legibility could be quantified, and future questions about word legibility could be answered by research on individual characters.

On the basis of these considerations, the present experiment was designed to determine the legibility of words as a function of horizontal and vertical resolution and character size, much as individual character legibility was determined by Erdmann and Neal (1966). Since other legibility studies (including those cited) suggest the need for more precise characterization of words, two potentially important variables—word familiarity and word length—were introduced as additional experimental parameters.

METHOD

The approach taken was to (*a*) design a test document which incorporated certain differences (in character height, word length, and word familiarity); (*b*) reproduce the document, adding systematic differences in vertical and horizontal resolution; (*c*) submit the resulting stimulus material to *Ss* to measure legibility; and (*d*) relate the legibility data thus derived to the physical variables under investigation. The procedure followed made it possible to start with a relatively small number of words, include them all in the original document, and then rearrange the material systematically after printing to prevent interaction between the content and the experimental parameters (word length and familiarity).

The content includes 144 familiar and 144 unfamiliar words selected from Thorndike and Lorge (1944) to represent a certain range of word lengths (see Figure 1). The "familiar" are among the 1000 words most common in English; the "unfamiliar" appear less often than once in a million words, and more often than once in 4 million. The row of individual characters below each set of words is made up of 17 letters in random order, with a, e, and o repeated. The whole document is in lowercase Dual Gothic type, selected for its relatively uniform stroke width and its simplicity.

Where the criterion measure is legibility, defined as the percentage of correctly identified letters or words, degradation variables are often introduced to control some aspect of the experimental situation—viewing time or distance, for example, or resolution,

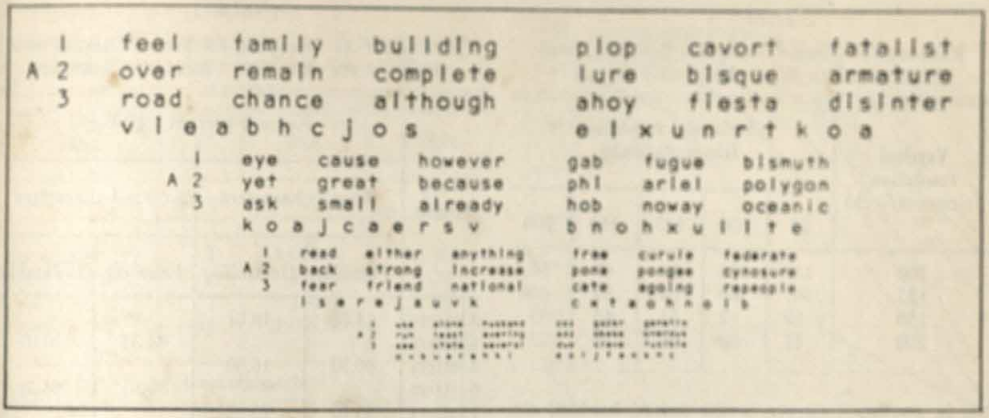


FIG. 1. Test document for word legibility study (one of four sessions).

size, or contrast in the stimulus material. In the work reported here, character size and resolution were varied, but high contrast was maintained throughout the printing process. In Sections B, C, and D of the test document, character heights were 40, 60, and 80 mils, respectively. In Section A (shown in Figure 1), characters were 100 mils in height, corresponding approximately to the 12-point type familiar in typescript and elsewhere. (The point system is standard in typography, but does not specify character size exactly. Character height is the better basis for direct comparisons of the effects of character size.)

The test document, mounted on the drum of a digital facsimile device for scanning and printing, was reproduced at 20 different combinations of horizontal and vertical resolution. The device scanned across the document in narrow bands, sampling periodically to determine if the document at successive points was black or white. The smaller the size of the scanning aperture and the more frequent the sampling, the greater the resolution. The resolution obtained can also be described, as it is more commonly, by the number of samples or elements per inch on the output copy. One can consider the scanned document to be covered by a grid, with each square in the grid representing a resolution element and the dimensions of the grid being the shape of the scanning aperture. Figure 2a shows how such a grid might reproduce one character on a document page. The character outlined represents the printed copy being scanned, and the shaded blocks show how the character would be "reported" by five successive scan lines. Figure 2b is an actual photograph of a character before and after digital facsimile reproduction.

Five different horizontal resolutions and four different vertical resolutions were used in this study. The size of the scanned element (square of the grid) for each of the 20 combinations is shown in Table 1. From the 20 copies reproduced, four test booklets were assembled so that each group of words (Section A, B, C, or D on the master document) was presented on five consecutive pages at the lowest vertical resolution in combination with each of the five (increasing) horizontal resolutions. The next five pages represented the next larger vertical resolution at each horizontal resolution (for the same group of words) and so on. Sections A, B, C, and D were presented in that order in the first test book-

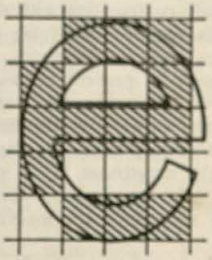


FIG. 2a. Digital facsimile "grid" showing black/white decision for each element area.

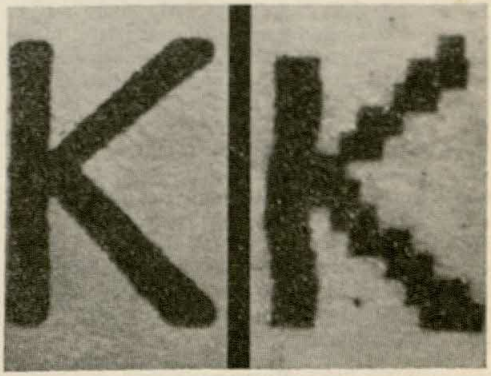


FIG. 2b. Printed character before and after digital facsimile reproduction at a resolution of 150 × 100 elements per inch (20 × Magnification).

TABLE 1

RESOLUTION COMBINATIONS (IN SQUARE MILS) COMPARED IN WORD LEGIBILITY STUDY

Vertical resolution (elements/inch)	Horizontal resolution (elements/inch)				
	75	100	125	150	200
100	133	100	80	67	50
125	107	80	64	53	40
150	89	67	53	44	33
200	67	50	40	33	23

let, in another sequence in the second booklet, etc. (Four of the 24 possible sequences of word groups were selected at random.)

The procedure permitted subsequent evaluation of the effects of 20 element areas and independent evaluation of the effects of horizontal and vertical resolution. Since the page order within each booklet was in the order of increasing resolution, the *S* would be expected to identify a larger percentage of words on each successive page.

The same words were read at all five horizontal resolutions while vertical resolution was held constant to eliminate confounding with word form and familiarity. This required modifying the method of limits (by using only an increasing series of horizontal resolutions) to prevent memorization of the stimulus material from an earlier page. This design was also practical in view of the limited time and resources available to produce the stimulus materials and conduct the experiment. Pilot experimentation prior to the main experiment indicated negligible practice effects when only the ascending series was used.

The *S*s for the experiment were 72 college students with 20/20 vision, natural or corrected, who volunteered for the task and received a fixed compensation. Each *S* was tested separately in a well-lighted booth. He was told that the object of the study was to determine how printing resolution was related to the legibility of letters and words. He was then asked to read aloud as many of the words and letters as possible in one of the four test booklets, page by page from front to back. If he could not pronounce a word, he was allowed to spell it, but *the word was not considered identified unless every letter was named correctly* (i.e., each word was considered as a unit). Proceeding at his own pace, each *S* spent approximately 1 hr. at the task. The "answers" recorded on coded sheets by the *E* were subsequently transferred to punched cards, and legibility results were calculated by means of programs written for the IBM 1620 computer.

RESULTS

From the accumulated data, the computer calculated the percentage of correctly identified words for each character size as a function of word length, familiarity, and vertical and horizontal resolution. Most of the 20 resolution combinations produced 100% legibility for words with 100-mil and 80-mil characters. The ranges were 93–100% and 70–100%, respectively. For 60-mil words the lower limit dropped sharply, the range being 30–100%; for 40-mil words, legibility ranged 0–100%.

TABLE 2

LEGIBILITY AS AFFECTED BY WORD LENGTH AND FAMILIARITY FOR TWO CHARACTER HEIGHTS

Word length	Average percent legibility			
	40-mil characters		60-mil characters	
	Familiar	Unfamiliar	Familiar	Unfamiliar
3 letters	43.66	18.01		
4 letters			84.31	70.07
5 letters	40.83	16.80		
6 letters			89.81	64.21
7 letters	43.03	25.62		
8 letters			91.46	77.48

This preliminary process of elimination made it apparent that the more detailed analysis should concentrate on the shorter characters, since the range of legibility for the taller was so limited. An analysis of variance was therefore performed for the 40-mil and 60-mil words to determine the effects (if any) of word length and familiarity. The results, summarized in Table 2, show statistically significant differences between familiar and unfamiliar words. For word length, on the other hand, the *F* ratio indicated negligible effects (see Table 3). The data were therefore combined across all word lengths for the remainder of the analysis.

The three remaining categories—letters, familiar words, and unfamiliar words—were compared as shown in Figure 3, with vertical and horizontal resolution combined to make element area the unit of resolution. (Subsequent interpretations of the data took into account the possible effect of the proportions of that unit, and considered vertical and horizontal resolution separately.) As in the

TABLE 3

ANALYSIS OF VARIANCE DUE TO WORD LENGTH AND FAMILIARITY FOR TWO CHARACTER HEIGHTS

Item	*df*	*MS*	*F* ratio
40-mil characters			
Familiarity (F)	1	15,000.59	14.46*
Word length (L)	2	310.85	.30
F × L	2	190.61	.18
Error	114	1,037.56	—
60-mil characters			
Familiarity (F)	1	9,655.29	15.54*
Word length (L)	2	723.93	1.17
F × L	2	440.99	.71
Error	114	621.38	—

Note.—Critical ratio for $F_{1/100} = 6.90$; for $F_{2/100}$ it is 4.82.
* $p < .01$.

Erdmann and Neal (1966) study of individual characters, legibility systematically increased with character height and with resolution (i.e., with smaller element areas). As to the relative legibility of words and letters, and the effect of familiarity, the results showed familiar words and individual letters to be consistently more legible than unfamiliar words, at all character heights.

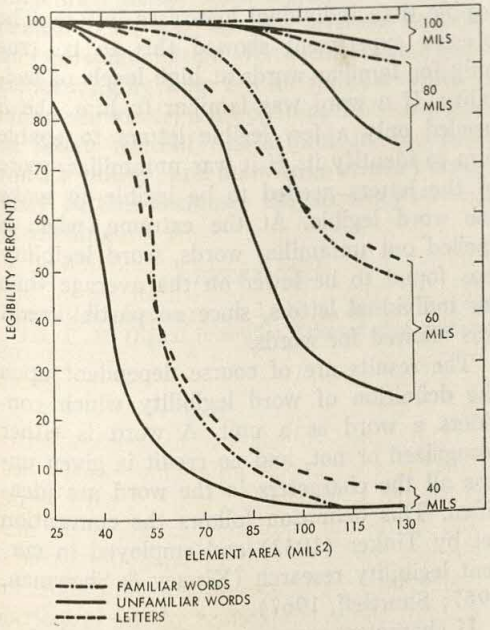


FIG. 3. Legibility of familiar and unfamiliar words, and of letters, at four character heights.

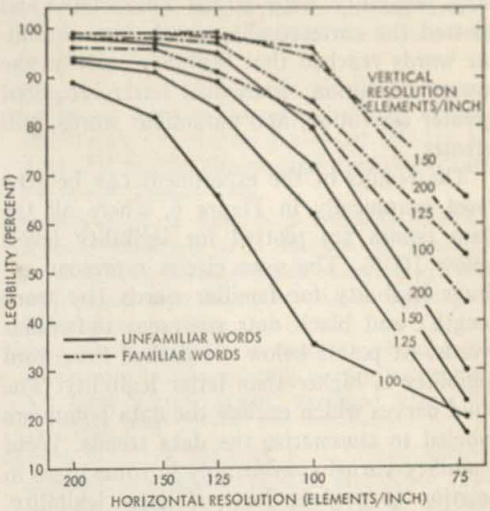


FIG. 4. Legibility of 60-mil characters, showing separate effects of horizontal and vertical resolution.

The independent effects of vertical and horizontal resolution were plotted in a representative graph for one character height (60 mils). As that graph (Figure 4) shows, the legibility of both familiar and unfamiliar words increased steadily with improved resolution in either direction. Again, familiar words were consistently the more legible. Another approach (see Figure 5) assumed a goal of

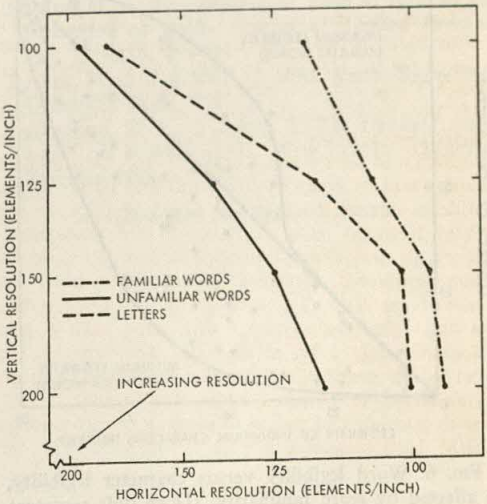


FIG. 5. Horizontal and vertical resolution required for 90% legibility with 60-mil characters.

90% legibility with 60-mil characters, and plotted the corresponding resolutions. Familiar words reached that legibility level at the lowest resolution; individual letters required greater resolution, and unfamiliar words, still greater.

The results of the experiment can be seen most graphically in Figure 6, where all the data points are plotted for legibility levels below 100%. The open circles represent average legibility for familiar words (by word length) and black dots represent unfamiliar words. At points below the dashed line, word legibility is higher than letter legibility. The solid curves which enclose the data points are labeled to summarize the data trends. Word legibility varied considerably in some cases in relation to a given level of letter legibility. The extreme case—a range of more than 50%—was observed at about 40% letter legibility. Word legibility varies less, of course, near the 0 and 100% levels of letter legibility. The product-moment correlations for the relationship between letter and word legibility are: .98 for familiar words, .94 for unfamiliar words, and .93 for words regardless of degree of familiarity.

These correlations depend in part upon the particular experimental conditions. If data had been collected only at points where the legibility was 0 or 100%, the correlations would, of course, be unity. In practice, however, the data points for the correlations show variability at each legibility level. Thus one would expect that for any reasonable distribution of conditions the trends indicated in this study would be verified; namely, the legibility of individual letters is more highly correlated with the legibility of familiar words than with that of unfamiliar words.

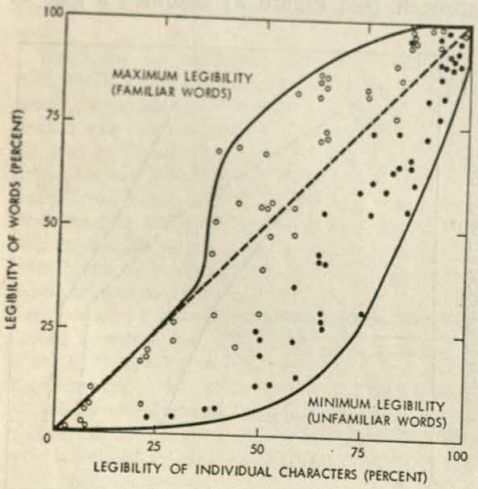


FIG. 6. Word legibility versus character legibility, as affected by word familiarity. (Open dots represent average legibility for familiar words; closed dots, for unfamiliar words.)

DISCUSSION AND CONCLUSIONS

Up to a point, the word legibility experiment serves to corroborate the findings of the related character legibility study (Erdmann & Neal, 1966). The results are comparable with respect to the effect of vertical and horizontal resolution on legibility for selected character heights. As expected, the taller the characters in use, the more legible the word at any given resolution. Correspondingly, for a given character height, word legibility increased with increasing resolution.

The specific purpose of the present study was served by the results which relate word legibility to character legibility, in the interest of predicting the former from the latter (which may be more readily measured). Previous work (Tinker, 1932, 1963) has led to the conclusion that words in general are more legible than individual lowercase letters. The present experiment showed this to be true only for familiar words at high levels of legibility. If a word was familiar to him, the *S* needed only a few legible letters to enable him to identify it. If it was unfamiliar, more of the letters needed to be legible to make the word legible. At the extreme, when *S* spelled out unfamiliar words, word legibility was found to be lower on the average than for individual letters, since no partial credit was allowed for words.

The results are of course dependent upon the definition of word legibility which considers a word as a unit. A word is either recognized or not, and no credit is given unless all the characters in the word are identified. This definition follows the convention set by Tinker (1932) and employed in current legibility research (Kinney & Showman, 1967; Shurtleff, 1967).

If characters were grouped into words but credit was given for individual characters, one

would expect familiar words to score somewhat higher than these groups of random characters. Longer familiar words might score higher than shorter words up to some limit, beyond which the identification of a few letters no longer insured identification of the word (even though length was not important when words were scored as a unit).

With the same scoring scheme (i.e., credit for characters), unfamiliar words should be approximately equal to random letters in legibility. Very unfamiliar words would probably require the identification of every character (except where knowledge of the common sequences of letters in English words offered an advantage). The important thing to note is that even where the measurement process differs, the legibility of words remains a function of word familiarity.

In conclusion, word familiarity was found to have considerable influence on unit word legibility, whereas word length had little effect. For familiar material, the average legibility of words can be expected to be equal to or higher than the average legibility of individual letters. However, for unfamiliar material (one might think of words in a foreign language, or technical information outside one's own field), word legibility can be expected to be less than the average legibility of individual letters. With correlation coefficients ranging from .98 for familiar words to .93 for all words, it is apparent that knowledge of letter legibility is helpful in predicting word legibility, and knowledge of word familiarity further increases the accuracy of such predictions.

REFERENCES

ARPS, R. B., ERDMANN, R. L., NEAL, A. S., & SCHLAEPFER, C. E. Digital facsimile: Transmission economy and copy quality. IBM ASDD Laboratory Memorandum ASDJ-M-097, Los Gatos, California, May 31, 1966.

CROOKS, M. N., ET AL. The effect of noise on the perception of forms in electrovisual display systems. Final report. Contract No. DA-49-007-MD-536. Institute for Applied Experimental Psychology, Tufts University, January 1959.

ERDMANN, R. L., & NEAL, A. S. Legibility of alphanumeric characters as a function of resolution parameters in digital facsimile output copy. IBM ASDD Laboratory Report 16.135, Los Gatos, California, July 14, 1966.

HOWES, D. H. On the interpretation of word frequency as a variable affecting speed of recognition. AD 32066 ASTIA, WADC Technical Report 54-282, June 1954.

HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word probability. *Journal of Experimental Psychology*, 1951, **41**, 401–410.

KINNEY, G. C., & SHOWMAN, D. J. The relative legibility of uppercase and lowercase typewritten words. *Information Display*, 1967, **4**, 34–39.

SALTZ, E. Thorndike-Lorge frequency and *m* of stimuli as separate factors in paired-associates learning. *Journal of Experimental Psychology*, 1967, **73**, 473–478.

SHURTLEFF, D. A. Studies in television legibility—a review of the literature. *Information Display*, 1967, **4**, 40–45.

SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition threshold for words. *Journal of Experimental Psychology*, 1952, **43**, 195–201.

TAYLOR, C. P. The relative legibility of black and white print. *Journal of Educational Psychology*, 1934, **25**, 561–578.

TINKER, M. A. The influence of form of type on the perception of words. *Journal of Applied Psychology*, 1932, **16**, 167–174.

TINKER, M. A. *Legibility of print*. Ames: Iowa State University Press, 1963.

THORNDIKE, E. L., & LORGE, I. *The teachers' word book of 30,000 words*. New York: Columbia Teachers College, 1944.

(Received November 2, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 5, 410–413

EFFECT OF LANE-CLOSURE SIGNALS UPON DRIVER DECISION MAKING AND TRAFFIC FLOW [1]

KENNETH PERCHONOK [2] AND PAUL M. HURST

Institute For Research, State College, Pennsylvania

A decision-theoretic model was applied to driver behavior on an urban expressway. The model permits inferential measurement of "responsiveness versus confusion" and "risk-taking predisposition." These variables, as well as 2 direct measures ("risk taking" and "hazard"), were measured in a field study of forced merging from a blocked lane. Resultant indexes were compared under 2 experimentally manipulated methods of lane closure: signal closure (SC) and conventional closure (CC). The SC method, which provides earlier warning of the closure, was superior to CC in terms of lower hazard and greater responsiveness versus confusion. There were also improvements in several traffic-flow characteristics.

The purpose of this study was to operationally define variables immediately involved in, and resulting from, the decision-making processes of drivers in a merging situation, to measure the changes in these variables under differential traffic-control systems, and hence to use these measured changes to evaluate those systems. The investigation proceeded in the context of stimulus-dependent response elicitation mediated by a decision-making process.

Briefly, the model for decision making states that a gap of size g will be accepted if it is larger than a driver's threshold, G, the smallest acceptable gap. It follows that $P_g(A)$, the probability of accepting a gap of size g, is equal to the probability, $P(g > G)$, that g is above threshold, and that, as g grows larger, $P_g(A) = P(g > G)$ increases from 0 to 1. It has been argued by Hurst (1964) that as drivers become better able to perceive and interpret environmental information their evaluations of incurring risk in gap acceptance/rejection become better, and their behavior becomes more consistent with respect to the immediate environment, and that this consistency can be measured by η, the correlation ratio between gap "size" and acceptance/rejection. Perchonok (1964) showed that this increased consistency is essentially equivalent to an increase in the slope of $P_g(A)$ for $0 < P_g(A) < 1$. Conversely, as drivers become less responsive to the relevant environment (they behave in a more confused manner), the response curve flattens out and the value of η is diminished. Thus η, as an index of "responsiveness versus confusion," measures the ability of drivers to cope with their respective environments.

In addition to changes in the degree of responsiveness, driver behavior also reflects the amount of risk that drivers are willing to tolerate. That is, if G, the gap-acceptance threshold, is lowered, $P(g > G)$ increases and $P_g(A)$, the probability of acceptance, increases. Thus, the gap-acceptance threshold reflects "risk-taking predisposition." The effect of change in driver responsiveness and risk-taking predisposition is illustrated for a hypothesized example in Figure 1, where, in comparison to Curve A, B shows decreased responsiveness and C shows a lowered threshold or increased risk-taking predisposition.

It can be seen that either type of change in the response curve is likely to reflect the probability of accepting a small gap, which is taken as our definition of "risk taking." That is, not only would a decreased threshold yield an increased probability of accepting a small gap, but decreased responsiveness would also tend toward that outcome.

Finally, a fourth measure was employed as an index of resultant "hazard." The measure

[1] This investigation was supported by Public Health Service Research Grant No. AC-66, from the Division of Accident Prevention. The authors wish to thank Edward Gervais and Frank DeRose, Jr., and their associates on the John C. Lodge Freeway Traffic Surveillance and Control Research Project, without whose cooperation this study could not have been done.

[2] Now at Cornell Aeronautical Laboratories, Buffalo, New York.

is given simply by the number of small gaps accepted; the greater the number of small gaps accepted, the greater the net hazard to drivers of merging vehicles and to those in the adjacent lane. It should be noted that although hazard is a function of risk taking, it is also a function of the distribution of available gap sizes—if very few small gaps are presented, hazard must be low even if the probability of accepting a small gap, when presented, is high.

In order to pursue this work most directly, it was decided to establish the "most relevant stimulus" with respect to gap-acceptance behavior. Such a measure of gap size was selected on empirical grounds from a list of 27 algebraic combinations of variables describing the kinematics of the merging vehicle and the vehicles forming the gap (Hurst, Perchonok, & Seguin, 1968).

METHOD

Data Collection

Data were recorded on 35-mm. film using an elevated camera overlooking a three-lane section of the John C. Lodge Freeway in Detroit, Michigan. The observation area was approximately 350 ft. long; the lane closure was established by a maintenance vehicle located in the medial lane at the downstream end of that area. Data collection started at 2:15 P.M. on two successive afternoons, one for each type of closure system.

Under one system, conventional closure (CC), the only warning given to drivers was supplied by the truck and a wooden panel, supporting flashing lights, attached to the rear of the truck. Under signal closure (SC) additional warning was provided by an overhead sign approximately 1,200 ft. upstream of the truck. This sign displayed electrically lighted signals consisting of a red "X" for the closed lane and a green arrow for each of the open lanes. In addition, a traffic engineer operated a suggested speed indicator, at the same location, following the same procedures used in any normally occurring lane closure.

The data reduction involved the transformation of pictorial information for each ¼ sec. into quantitative form via the Pictorial Data Transducer system developed by the Institute For Research, State College, Pennsylvania. This system involves a semi-automated film-scanning device, from which vehicular positional data on IBM cards are obtained, and a battery of computer programs to organize this output into an analyzable format. A more complete description of the system is given by Seguin and Perchonok (1965); the details of the transducer itself are discussed by Hicks (1964).

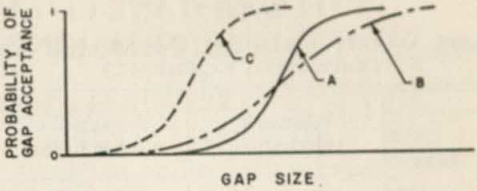


FIG. 1. Three hypothetical response curves illustrating differential responsiveness and risk-taking predisposition.

Definitions and Formulas

"Gap size" was measured by a decision point which was defined as that time, to the nearest ¼ sec., at which the longitudinal locations of the rear of the lead vehicle defining the gap and the front of the subject vehicle in the medial (closed) lane were coincident. For the purposes of this study only those gap occurrences were included for which the lead vehicle had a higher velocity than the subject vehicle. Gap size as determined by Hurst et al. (1968) was measured as:

$$\text{gap size} = [H_F/V_F] - T_{F\text{-}S}$$

where H_F = the distance (ft.) between the lead and following vehicles, minus the length of the subject vehicle; V_F = speed (ft/sec) of the following vehicle; and $T_{F\text{-}S}$ = the minimum time (sec.) required by the following vehicle to match the velocity of the subject vehicle assuming maximum deceleration with a coefficient of friction of .6, and a response time of .75 sec.

"Responsiveness versus confusion" was measured by η, the correlation ratio, which is the square root of the proportion of acceptance/rejection behavior attributable to gap size. A higher value of η implies greater responsiveness and, therefore, less confusion.

"Risk-taking predisposition" was measured by $\bar{G}$, the mean value of the gap-acceptance threshold as computed by transforming the acceptance curve to a frequency distribution of thresholds and computing the mean as suggested by Spearman (Guilford, 1954).

A "small gap" was defined as any gap whose size was less than the $\bar{G}$ computed from the pooled data. It was, then, any gap which the average driver would consider unacceptable.

"Risk taking" was defined as the number of small gaps accepted divided by the number of small gaps presented. It was, thus, an estimate of the probability that a small gap would be accepted.

"Hazard" was defined as the number of small gaps accepted.

RESULTS

Volumes under the two closure systems were essentially equivalent: under signal closure (SC) the volume was 67.1 vehicles/min, while under conventional closure (CC) the volume was 67.7 vehicles/min. However, relative

TABLE 1

LANE VOLUMES FOR SIGNAL CLOSURE (SC) AND CONVENTIONAL CLOSURE (CC)

Lane	Volumes in vehicles/min		Proportion of volume by lane	
	SC	CC	SC	CC
Shoulder	31.8	30.5	.47	.45
Center	30.6	25.8	.46	.38
Medial	4.7	11.4	.07	.17
Total	67.1	67.7	1.00	1.00

lane volumes did change as a function of closure (Table 1). Table 1 further explicates the change in relative lane volumes in terms of the probabilities of entering the observation area in each of the three lanes.

Chi-square tests showed the probability that a vehicle entering the observation area did so in the medial lane was significantly ($p < .002$) lower under SC and that the probability of entering in the shoulder lane was not significantly changed. Of the vehicles not entering the medial lane, the likelihood of entering the center, as opposed to shoulder, lane increased under SC but not significantly ($p = .17$). The relative lane volumes leaving the observation area were quite similar under the two closure systems.

Other changes in traffic characteristics that were noted are shown in Table 2. The probabilities of these results occurring if chance alone had been in operation were sufficiently small to allow rejection of the hypotheses of "no difference."

The curves in Figure 2 are the empirical response curves under the two systems of closure. The two vertical lines represent the respective $\overline{G}$s.

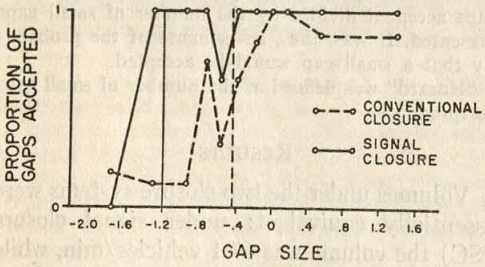


FIG. 2. Proportion of gap acceptances as a function of gap size for two types of closure.

TABLE 2

THREE TRAFFIC-FLOW CHARACTERISTICS FOR SIGNAL CLOSURE (SC) AND CONVENTIONAL CLOSURE (CC)

Characteristic	SC	CC	p
Median speed of medial lane vehicles at decision point (ft/sec)	18.7	14.0	.058[a]
Median number of vehicles in observation area	16.0	17.0	.042[a]
Median waiting time from area entrance to gap acceptance (sec.)	6.0	9.0	.012[b]

[a] Significance level using one-tailed Mann–Whitney U test that observations under one system were stochastically higher than those under the other system (Siegel, 1956).
[b] Two-tailed test.

The values of the decision-making correlates for SC and CC along with significance levels for respective changes are given in Table 3. All differences, except that for risk taking, had small probabilities of chance occurrence and were therefore accepted as "real" differences.

DISCUSSION AND CONCLUSIONS

The similarity of overall volumes seems to indicate that observed differences are reasonably attributable to the type of closure, as opposed to the nature of the traffic. The relative effect of SC upon lane volumes was as expected: the early warning of lane closure was successful in inducing drivers to leave the medial lane causing a concomitant increase of volume in the center lane. Thus, the signal system was successful with respect to its primary function.

Additional benefits included a decrease in the slowing of subject vehicles and a shorter waiting time before leaving the closed lane. The significant reduction of vehicle density in the observation area implies that vehicle velocities tended to be higher, and that potential capacity was increased. Thus, the effectiveness of the SC apparently extended to a general improvement in traffic flow.

The decreased confusion during SC is graphically exhibited in Figure 2 which exhibits a tendency toward more appropriate behavior than under CC; namely, the likelihood of accepting very small gaps and of rejecting large gaps is lower under SC than CC. It should be noted that this phenomenon obtains even in the face of a changed thresh-

TABLE 3

DECISION-MAKING CORRELATES FOR SIGNAL CLOSURE (SC) AND CONVENTIONAL CLOSURE (CC)

Variable	SC	CC	Test	p
Responsiveness versus confusion (η)	.890	.714	Wilcoxon[a]	.063
Risk-taking predisposition ($\bar{G}$)	−1.204	−.428	Wilcoxon[b]	.024
Risk taking $\left(\frac{\text{no. small gaps accepted}}{\text{no. small gaps}}\right)$	.43	.30	Fisher[b]	.808
Hazard (no. small gaps accepted)	3	9	Chi-square[a]	.050

[a] The Wilcoxon matched-pairs signed-rank test was used to detect stochastic differences. The directions of change in "responsiveness versus confusion" and in "hazard" were predicted; thus one-tailed hypotheses were tested.
[b] Two-tailed.

old. It can be inferred that although drivers under SC exhibited a greater risk-taking predisposition—a willingness to accept smaller gaps—their lower level of "confusion" permitted them to better behave in correspondence with the environment, with the result that there were no significant differences in actual risk taking by individual drivers. Thus, the effect of SC in reducing "confusion" paid off. The early alerting to lane closure apparently allowed drivers to better assimilate information about the environment and thus to respond more appropriately to it.

The increased risk-taking predisposition under SC was unexpected. An explanation, although ad hoc, is available. It had been observed via the television surveillance system in Detroit that some drivers ignore the lane-closure signal and remain in the closed lane, taking advantage of its reduced volume. Since drivers receive an earlier warning under SC, it seems appropriate to assume that of the drivers in the observation area a greater proportion were there by choice under SC as compared to CC. Thus, instead of interpreting the increased predisposition toward risk taking as a product of the situation in which the driver finds himself, it seems likely that the early warning acts as a selection device encouraging the less aggressive driver to leave the closed lane. Thus, the more aggressive driver remained to take advantage of the reduced medial lane volume at the possible cost of encountering higher risk when, as he ultimately must, he changes lanes.

Finally, it can be seen that measured "hazard" was greater under CC than SC. The differential comparison between "risk taking" and "hazard" is attributable to the reduced medial lane volume under SC. Here, again, the success of the signal system is obvious.

It may be concluded that the SC system provided a significant safety improvement over the CC system in terms of reduced objective "hazard" in gap acceptance. The means whereby this was achieved were a reduction in the proportion of vehicles entering the observation area in the closed lane and (by inference) a decrease in "confusion" by drivers of merging vehicles.

REFERENCES

GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.

HICKS, W. W., JR. A pictorial data transducer. Division of Highway Studies, Institute For Research, State College, Pa., Report No. 4, 1964.

HURST, P. M. Errors in driver risk-taking. Division of Highway Studies, Institute For Research, State College, Pa., Report No. 2, 1964.

HURST, P. M., PERCHONOK, K., & SEGUIN, E. L. Vehicle kinematics and gap acceptance. *Journal of Applied Psychology,* 1968, **52**, 321–324.

PERCHONOK, K. The measurement of driver errors. Division of Highway Studies, Institute For Research, State College, Pa., Report No. 1, 1964.

SEGUIN, E. L., & PERCHONOK, K. Vehicle interactions on an urban expressway. Division of Highway Studies, Institute For Research, State College, Pa., Report No. 6, 1965.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

(Received November 17, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 5, 414–422

LEADERSHIP PERFORMANCE OF NURSING SUPERVISORS AT TWO ORGANIZATIONAL LEVELS [1]

STANLEY M. NEALEY AND MILTON R. BLOOD [2]

University of Illinois

The relations of leader style and leader behavior to work-group performance and subordinate job satisfaction have been investigated for 22 1st-level and 8 2nd-level nursing supervisors in a Veterans Administration Hospital of 1680 beds. Task-oriented (low least preferred co-worker—LPC) leaders received higher performance ratings at the 1st level of supervision, while relationship-oriented (high LPC) leaders performed better at the 2nd level of supervision. Subordinate's job satisfaction was positively related to leader Consideration at both levels of supervision. However, Initiating Structure leader behavior contributed to high subordinate job satisfaction at the 1st level of supervision but low subordinate job satisfaction at the 2nd level. These differences in effective supervisory patterns point to differences in situational leadership demands at the 2 supervisory levels.

The present study was designed to explore in a field setting the relations of leadership style and leadership behavior on the one hand to measures of work-group performance and work-group satisfaction on the other, and to compare these relations across two organizational levels. In other words the aim is to discover if and how leadership demands differ by organizational level.

There are several reasons for hypothesizing that the first-level supervisor and his boss, the second-level supervisor, will face different leadership demands. For one thing, leadership theory and leadership research have for several decades been moving away from the idea of universal leadership traits and the "one best way" to lead. In place of these earlier oversimplifications, leadership effectiveness is now generally conceded to involve an interaction between the leader's characteristics, his behavior, the nature of the followers, and the characteristics of the particular leadership situation, including the nature of the task and the organizational setting. Terman (1904) initially suggested this approach, and studies by Carter, Haythorn, Shriver, and Lanzetta (1951), Hemphill (1949), Stogdill (1948), Flanagan (1949), and many others (see Gibb, 1954) have led to its general acceptance. The interactional conception of leadership is the basis for Fiedler's (1964, 1967) model of leadership effectiveness and figured prominently in reviews of research on supervision and productivity by Dubin (1965) and Sales (1966). The point is that a number of situational variables, for instance, follower characteristics, leader power, and group task, may vary markedly by organizational level. Leader behavior, in order to be effective, may need to vary accordingly.

A number of writers in industrial administration, among them Barnard (1950), Niles (1949), and Pfiffner and Sherwood (1960), have hypothesized that management functions do vary from level to level within organizations. Georgopoulos and Mann (1962) found that technical skills become less important as managers move upward in organizations. Argyris (1964) and Katz and Kahn (1966) have expressed similar views.

A partially contrasting perspective on the above question of differential demands by management level is provided by several studies in which managers at different organizational levels have expressed their opinions about what qualities lead to success (see

[1] This study was conducted under Contract NR 177-472, Nonr 1835(36) (Fred E. Fiedler, principal investigator). The authors wish to express their appreciation to Otto Schaefer, hospital director, and to Mark Ardis, chief of staff, for permission to conduct the study. They are indebted to Roy Hamlin, research psychologist, for liaison and critical reading of the manuscript, to Fred E. Fiedler for guidance and support, and to the entire nursing service for their help and cooperation throughout the study. The authors particularly wish to thank James Catney, chief of nursing, for his active participation at every stage of the study.

[2] Now at University of California, Berkeley.

Fleishman, 1953; Pellegrin & Coates, 1957; Porter & Henry, 1964). These studies all found some differences by level, but these differences were not striking. On the whole, the managers surveyed did not feel that effective patterns of leadership attributes differ greatly by organizational level. Related studies by Rosen (1961) and by Triandis (1959a, 1959b) reinforce this conclusion.

The opinions of industrial administration scholars and those of managers thus seem in only partial agreement that organizational level is an important variable in determining which pattern of leadership behavior will be most effective. Whatever may be the eventual empirical resolution of these differences, if we are willing to entertain the notion that the leadership demands on top managers may be different in kind as well as degree from the leadership demands at lower levels, then we must face the possibility that experience at lower levels may have minimal value, or may even be disadvantageous, to effective performance at higher levels in the organization.

METHOD

Research Site and Subjects

This study was conducted in a Veterans Administration hospital of 1,680 beds. The *Ss* were nursing personnel at four levels in the nursing service as follows: 248 nursing assistants, 43 staff nurses, 22 head nurses, and 8 unit supervising nurses. All *Ss* except nursing assistants were RNs.

The aspect of the study which is of concern here is a comparison of certain leadership variables across two levels of the nursing staff. For this purpose head nurses will be considered as *first-level supervisors* and will be referred to as such from here on. Each head nurse supervised directly three to eight nursing assistants as well as one or two staff nurses. The hospital was organized into six units, each with roughly equivalent functions. The nursing function in each unit was under the supervision of a unit supervising nurse who reported directly to the hospital's chief of nursing. An evening supervising nurse and a night supervising nurse had responsibility for the nursing function for the whole hospital during their respective shifts. For purposes of analysis, these two have been lumped with the six unit supervisors and these eight nurses will hereafter be referred to as *second-level supervisors.*

Predictor Variables

Two predictor variables have been used in the present study. The first was a measure of leadership style, specifically Fiedler's (1964) measure of the leader's esteem for his least preferred co-worker (LPC). The *S* is asked to think of the person with whom he has had the most difficulty in working. Then he describes this person by means of 17 evaluative bipolar adjective scales. A high LPC score indicates that *S* is relatively accepting of his least preferred co-worker (even though he was hard to work with). Fiedler (1967) hypothesized that *S* with a high LPC score is generally oriented toward good group relations when in a leadership role. A low LPC score indicates that *S* has low esteem for the person with whom he cannot work well. Fiedler (1967) hypothesized that *S* with a low LPC score is generally oriented toward task accomplishment when in a leadership role.

The second predictor variable was a measure of leader behavior as described by the leader's subordinates. Twenty-four items (the 12 highest loading and purest loading items on each dimension) were used from the Leadership Behavior Description Questionnaire (LBDQ), developed at Ohio State University (see Stogdill & Coons, 1957, pp. 108–109). It yields two scores. The first, Consideration, indicates the degree to which the leader exhibits a human-relations orientation toward his subordinates, while the second, Initiating Structure, indicates the degree to which the leader performs the management functions of planning, organizing, controlling, and pressing subordinates for production.

Criterion Variables

Two types of criterion variables were employed. The first type consisted of ratings of the performance of organizational subparts. These ratings were made by superiors one level above the supervisor of the subpart in question. Thus, the performance of the second-level supervisors was rated by the chief of nursing and by his two principal assistants. The performance of first-level supervisors was rated by second-level supervisors. It is important to note that these ratings involved the performance of the whole subpart of the organization rather than just the leader of that subpart. Interest was in organizational performance rather than leader ratings. These performance ratings were made on 4-point scales and involved three areas of nursing performance as follows: (1) physical patient care; (2) information about patients, the extent to which work-group members knew patients and their individual treatment schedules; and (3) human relations skill, skill in handling difficult emotional or social situations, understanding the why of patient behavior. A fourth rating, also on a 4-point scale, was a global rating of job performance. The 4 points on these scales were labeled, "much above average," "above average," "about average," and "a little below average" (among those subordinates rated).

A second type of criterion was the mean job satisfaction of the immediate subordinates of each supervisor. Thus the subordinate job-satisfaction score of second-level supervisors consisted of the mean job satisfaction of the head nurses and staff nurses who reported to them. Likewise the subordinate job-satisfaction score of a first-level supervisor

TABLE 1

CORRELATIONS BETWEEN LEADERS' LEAST PREFERRED CO-WORKER SCORE AND RATED PERFORMANCE CRITERIA AT TWO LEVELS OF SUPERVISION

Performance criteria	First-level supervisors ($N=21$) r_1	Second-level supervisors ($N=8$) r_2	$p(r_1-r_2)$
Physical patient care	−.395	.584	<.05
Information about patients	−.373	.646	<.05
Human relations	−.063	.817*	<.05
Global performance	−.219	.790*	<.05

Note.—Throughout the article, two-tailed p values have been used.
* $p<.05$.

consisted of the mean job satisfaction of the nursing assistants under her. Job satisfaction was measured by means of the Job Descriptive Index (JDI) developed at Cornell University (Smith, Kendall, & Hulin, 1968). This index yields satisfaction scores in five areas: the work itself, supervision, people, pay, and promotional opportunities.

Several other variables are included in the data analyses of this study. These include: (*a*) years of service as a registered nurse, (*b*) intelligence as measured by the Wonderlic Personnel Test (Wonderlic, 1959), (*c*) a rating of leadership power on which the chief of nursing rated second-level supervisors who in turn rated first-level supervisors, and (*d*) a measure of the leader's esteem for her immediate subordinates as a group. The leadership power rating consisted of nine items. Five items measured the supervisor's "expert power" in doing the work, for example, how well the supervisor knows the work relative to her subordinates, the amount of knowledge and skill the supervisor has, and her ability to give explicit and detailed guidance to subordinates. Four items measured the supervisor's "delegated power," for example, her degree of authority to reward or punish subordinates, her influence with superiors in having recommendations accepted, and her power to set work standards for subordinates. The expert- and delegated-power subscores were correlated .737 among first-level supervisors and .777 among second-level supervisors. These two scores have been combined as a measure of leader power in some of our analyses. As shall be seen later, this rating of the leader's power was highly correlated with the ratings of work-group performance in the present samples. The measure of leader's esteem for subordinates consisted of 10 evaluative bipolar adjective scales[3] plus five questions measuring the leader's confidence in and friendship for her subordinates and her evaluation of how well they do their work.

TABLE 2

CORRELATIONS BETWEEN LEADERS' LEAST PREFERRED CO-WORKER SCORE AND RATED LEADER POWER AT TWO LEVELS OF SUPERVISION

Rated leader power	First-level supervisors ($N=21$) r_1	Second-level supervisors ($N=8$) r_2	$p(r_1-r_2)$
Expert power	−.337	.471	*ns*
Delegated power	−.224	.766*	<.05
Total power	−.315	.614	<.05

* $p<.05$.

Data Analyses

Correlations between predictor variables and criterion variables will be reported separately for the two levels of supervision. Two-way analyses of variance were performed to determine if pairs of predictor variables interact to determine work-group performance and job satisfaction. One set of ANOVAs were performed using the leader's LPC score and the leader's esteem for subordinates as independent variables. Another set of ANOVAs involved the leader-behavior variables of Consideration and Initiating Structure as independent variables. These ANOVAs produced a number of significant main effects at both supervisory levels, but significant interactions occurred only at the first level of supervision. ANOVA results will be reported only for these significant interactions. In other words, correlations will be reported rather than ANOVA main effects.

RESULTS

LPC Scores as Predictors

The correlations between the leadership style measure (LPC) and the four ratings of performance by superiors are shown in Table 1. As can be seen the correlations differ in sign at the two supervisory levels. In general the high LPC (relations oriented) second-level supervisors received better ratings from superiors, while at the first level of supervision low LPC (task oriented) nurses were judged the better supervisors. However, none of the correlations at the first level of supervision quite reaches the two-tailed significance levels used for all significance tests throughout this paper. Column 3 of Table 1 shows that the correlations between LPC and performance were significantly different at the two levels of supervision.

[3] These 10 items are identical to Fiedler's (1967, Appendix A) "group atmosphere" scale. The part-whole correlation between the sum of these 10 items and the sum of the total 15 items was .958 among the first-level supervisors.

Ratings of the leadership power enjoyed by supervisors at both levels were made by superiors just as were the ratings of performance criteria. Raters were instructed to rate the expert and delegated power of the supervisor herself, but to rate the *whole work group* on the performance criteria. Despite this different emphasis, the power and criterion ratings turned out to be highly correlated at both supervisory levels. The power total score and the global rating of performance were correlated .884 and .856 ($p < .01$ in both cases) at the first and second levels, respectively, of supervision. Therefore, the power ratings are treated as another indication of the excellence of job performance as seen by superiors. Table 2 shows the correlations between leader LPC scores and rated leadership power at the two levels. Again high LPC (relation oriented) second-level supervisors receive good ratings, and again LPC predicts in different directions at the two levels.

The correlations between leadership style (LPC) and the job satisfaction of subordinates are shown in Table 3. The LPC score of supervisors did not correlate significantly with any area of subordinate job satisfaction or with the sum of the five job-satisfaction scores.

TABLE 3

CORRELATIONS BETWEEN LEADERS' LEAST PREFERRED CO-WORKER SCORE AND SUBORDINATE JOB SATISFACTION AT TWO LEVELS OF SUPERVISION

Job-satisfaction area	First-level supervisors ($N=21$) (r_1)	Second-level supervisors ($N=8$) (r_2)	$p(r_1-r_2)$
Work	−.273*	.439	ns
Supervision	−.275	.340	ns
People	−.288	−.407	ns
Pay	.208	.094	ns
Promotion	−.116	−.475	ns
Total satisfaction	−.199	.326	ns

* None of the correlations in this table is significant at the .05 level.

Relations between Predictors

Before moving on to the correlations between leader behavior, as perceived by subordinates, and the criterion variables, the relations among the two types of leadership predictor variables should be reported. Table 4 shows the correlations between the leadership style measure (LPC) and the leader-behavior dimensions Consideration and Initiating Structure. None of the correlations in Table 4 approaches significance. Apparently these two types of predictor variables were quite independent. Readers who may have fallen into the habit of interpreting high LPC scores as indicating high Consideration and low LPC scores as indicating high Initiating Structure should take note of this finding.

LBDQ Scores as Predictors

Descriptions of leader behavior as perceived by subordinates failed to predict ratings of performance by superiors. Table 5 shows the correlations between the LBDQ scores Initiating Structure and Consideration on the one hand and rated performance and rated power on the other. Only one of the correlations is significant. Ratings of expert power were correlated .494, $p < .05$, with Initiating Structure scores among first-level supervisors. One might conclude very cautiously that first-level supervisors rated high in Initiating Structure by their subordinates are well thought of by their superiors. Furthermore, the significance values shown in Column 7 of Table 5 lead to the conclusion that among first-level supervisors Initiating Structure behavior was valued more by superiors than was Consideration behavior.[4]

[4] These differences were tested by means of Hotelling's t test for the significance of the difference between r_{yx} and r_{xx} when computed from the same population. (See Walker & Lev, 1953, p. 256.)

TABLE 4

CORRELATIONS BETWEEN LEADERS' LEAST PREFERRED CO-WORKER SCORE AND LEADER BEHAVIOR AT TWO LEVELS OF SUPERVISION

Leader behavior dimension	First-level supervisors ($N=21$) r_1	Second-level supervisors ($N=8$) r_2	$p(r_1-r_2)$
Initiating Structure	−.167	.113	ns
Consideration	−.163	.418	ns

TABLE 5

CORRELATIONS BETWEEN LEADER BEHAVIOR AND SUPERIORS' RATINGS OF PERFORMANCE AND POWER AT TWO LEVELS OF SUPERVISION

Superiors' ratings	First-level supervisors ($N=22$)		Second-level supervisors ($N=8$)		Across levels		Within levels	
	Initiating Structure r_1	Consideration r_2	Initiating Structure r_3	Consideration r_4	$p(r_1-r_3)$	$p(r_2-r_4)$	$p(r_1-r_2)$	$p(r_3-r_4)$
Physical patient care	.107	.131	.150	.012	*ns*	*ns*	*ns*	*ns*
Information about patients	.240	.313	.090	.306	*ns*	*ns*	*ns*	*ns*
Human relations	.010	.103	.057	.302	*ns*	*ns*	*ns*	*ns*
Global performance	.413	.077	.232	−.015	*ns*	*ns*	<.05[a]	*ns*
Expert power	.494*	.162	−.144	.409	*ns*	*ns*	<.05	*ns*
Delegated power	.098	−.272	.201	.157	*ns*	*ns*	<.05	*ns*
Total power	.362	−.009	−.017	.333	*ns*	*ns*	<.05	*ns*

[a] This significance level may seem surprising since $r=.413$, $df=20$, is not significantly different from zero. The difference between $r=.413$ and $r=.077$ is significant because Consideration and Initiating Structure are correlated .704 in this sample.
* $p<.05$.

While leader behavior as rated by subordinates was only slightly correlated with performance ratings by superiors, it was closely related to subordinate job satisfaction. Table 6 shows the correlations between LBDQ scores and job satisfaction as measured by the JDI. At the first level of supervision both Consideration and Initiating Structure are positively related to all five aspects of subordinates' job satisfaction and to the job-satisfaction sum. All but two of these relations are significant beyond the .05 level. As Column 7 of Table 6 shows, Consideration was correlated with subordinate job satisfaction more highly than was Initiating Structure. Clearly, nursing assistants positively valued both Consideration and Initiating Structure behavior on the part of the first-level supervisor, but they valued the former more than the latter. At the second level of supervision the situation appeared somewhat different. Table 6 shows that Consideration was on the whole positively related to the job satisfaction of subordinates, but

TABLE 6

CORRELATIONS BETWEEN LEADER BEHAVIOR AND SUBORDINATE JOB SATISFACTION AT TWO LEVELS OF SUPERVISION

Job-satisfaction area	First-level supervisors ($N=22$)		Second-level supervisors ($N=8$)		Across levels		Within levels	
	Initiating Structure r_1	Consideration r_2	Initiating Structure r_3	Consideration r_4	$p(r_1-r_3)$	$p(r_2-r_4)$	$p(r_1-r_2)$	$p(r_3-r_4)$
Work	.520*	.762**	.087	.306	*ns*	*ns*	<.05	*ns*
Supervision	.557**	.790**	−.712*	.820*	<.01	*ns*	<.05	<.05
People	.657**	.716**	.494	−.144	*ns*	<.05	*ns*	*ns*
Pay	.256	.524*	−.537	.263	*ns*	*ns*	*ns*	*ns*
Promotion	.348	.622**	−.669	.289	<.05	*ns*	*ns*	*ns*
Total satisfaction	.501*	.809**	−.170	.624	*ns*	*ns*	<.01	*ns*

* $p<.05$.
** $p<.01$.

that Initiating Structure scores of second-level supervisors tended to be negatively related to the job satisfaction of their subordinates. The correlations between the leader-behavior dimensions and satisfaction with supervision were not only significantly different from zero, in opposite directions, but were significantly different from one another ($p < .05$). In this situation it appeared the first-level supervisors wanted their own superior to evidence Consideration but *not* Initiating Structure behavior. This is in contrast to the first-level supervisor's own subordinates, the nursing assistants, who valued *both* Consideration and Initiating Structure.

The above findings involving LBDQ scores should be interpreted in the light of the relations between the Consideration and Initiating Structure dimensions themselves. Among second-level supervisors these two dimensions were correlated −.570, while at the first level of supervision they were correlated +.704, $p < .01$. These two dimensions of the LBDQ are, of course, designed to be orthogonal. The 12 items in each scale were selected because of their orthogonal factor loadings based on an industrial population as reported by Stogdill and Coons (1957, pp. 108–109). While Stogdill and Coons report (p. 110) that their dimension scores intercorrelated only −.02 in an additional sample of 122 foremen, the results of this study indicate a much lower degree of factor independence. As a check the LBDQ scores from the total sample of 251 *Ss* who had completed the LBDQ were factor analyzed. The principal axis factor analysis even without rotation strongly supported the two-dimensional factor structure and a priori item scoring of the present data. Nevertheless, the high intercorrelations between the Consideration and Initiating Structure item composites in the samples of this study and the fact that these intercorrelations differed in direction from one organization level to another should be kept in mind. The authors would further recommend that the factor independence of the LBDQ dimensions be checked as a matter of course in future studies using this instrument, and that these checks be carried out separately for groups of *Ss* in separate situations or experimental treatments.

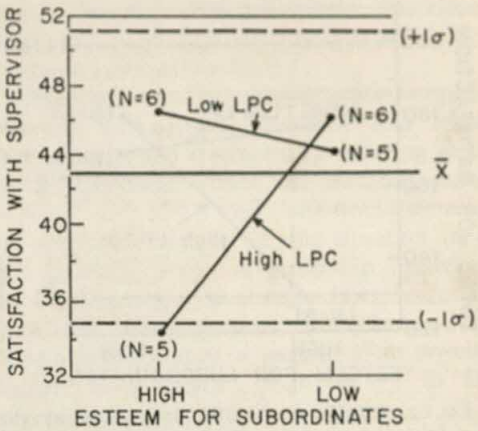


FIG. 1. Significant ANOVA interaction at the first level of supervision. The leaders' least preferred coworker (LPC) score and esteem for subordinates interact to determine satisfaction with supervision.

As previously noted, two-way analyses of variance were performed to see if the leadership predictor variables would interact significantly to yield differences on the criterion variables. For the most part they did not, but at the first level of supervision there occurred a set of significant interactions that are rather revealing. These analyses featured as independent variables the leadership style measure (LPC) and the leader's esteem for her subordinates (these two variables were correlated −.008, $N = 21$).

The first-level leader's LPC score and her esteem for her subordinates interacted to produce differences in her subordinates' satisfaction with supervision ($F = 5.255$, $df = 1/18$, $p < .05$) and total job satisfaction ($F = 6.992$, $df = 1/18$, $p < .05$). These interactions are charted in Figures 1 and 2. As can be seen, first-level supervisors with high LPC scores and high esteem for their subordinates stand out as having highly dissatisfied subordinates. This finding was partially explained by the fact that the same two independent variables (LPC and esteem for subordinates) interacted to produce significant differences in the Consideration ratings of these first-level supervisors ($F = 5.166$, $df = 1/18$, $p < .05$). This interaction is charted in Figure 3. Nurses with high LPC scores and high esteem for subordinates were rated by those subordinates as lowest in Consideration

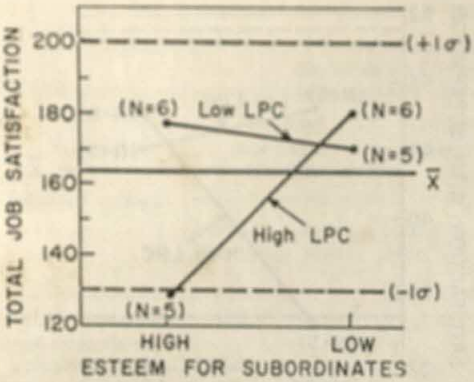


FIG. 2. Significant ANOVA interaction at the first level of supervision. The leaders' least preferred co-worker (LPC) score and esteem for subordinates interact to determine total job satisfaction.

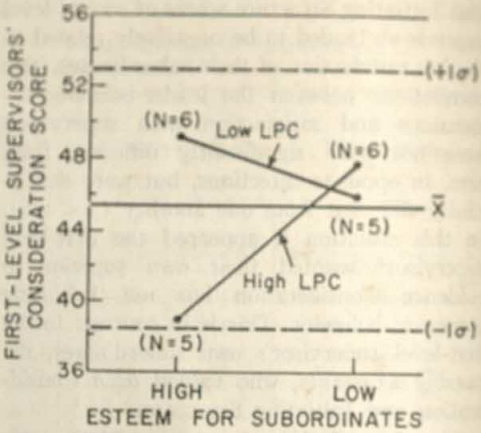


FIG. 3. Significant ANOVA interaction at the first level of supervision. The leaders' least preferred co-worker (LPC) score and esteem for subordinates interact to determine Consideration.

behavior. Although the interaction did not reach significance, this same group of nurses was also lowest in Initiating Structure behavior. In other words, these high LPC first-level supervisors who like and trusted their subordinates were seen by them as generally "inactive" in leadership behavior. Their subordinates in turn were highly dissatisfied with their supervision and with their total jobs.

Relations between Criteria

At each of the two supervisory levels the correlations were computed between the five aspects of job satisfaction of subordinates and the four ratings of performance made by superiors. None of the 20 correlations at either level reached statistical significance. It is interesting to note, however, that 18 out of 20 of the correlations at the first level of supervision are positive, while 13 out of 20 are negative at the second level of supervision. A χ^2 test of the frequency of plus and minus correlations at the two supervisory levels yielded $\chi^2 = 12.90$, $df = 1$, $p < .001$. It should be pointed out that the 40 correlations on which this χ^2 is based are not independent since there are substantial positive correlations within the JDI subscores and within the performance rating areas. The resulting χ^2 is thus difficult to interpret. Conservatism compels us to stop short of the tempting conclusion that the role expectations sent from superior and subordinate are compatible in the case of the first-level supervisor but contradictory in the case of the second-level supervisor; and that here, contrary to the usual situation, the second-level supervisor is really the "man in the middle" caught between conflicting demands from superiors and subordinates.

DISCUSSION

The present study was based on the single simple hypothesis that effective leadership style and leadership behavior patterns would differ across supervisory levels. This hypothesis was confirmed. In general, good performance ratings by superiors went to high LPC (relations oriented) second-level supervisors and to low LPC (task oriented) first-level supervisors (see Tables 1 and 2). Taking subordinate job satisfaction as a leader goal, it was found that Consideration behavior on the part of the leader was positively valued by subordinates at both supervisory levels. Initiating Structure behavior was also positively valued by subordinates at the first level of supervision, but was negatively valued at the second level of supervision.

A post hoc interpretation of the job-satisfaction findings comes readily to hand. The subordinates of second-level supervisors were RNs and shared the same professional qualifications as their leader. They liked her to be considerate but they did not want her to structure their jobs for them. On the other

hand, the subordinates of the first-level supervisors looked to them for *both* consideration and structure because they did not have the professional qualifications to handle psychiatric patients alone. Their leader was professionally qualified and they looked to her for structure. This suggests that the amount of Initiating Structure behavior a subordinate welcomes from a superior is partly a function of the difference between them in the degree of task expertise they possess.

Since Initiating Structure and Consideration were highly positively correlated at the first level of supervision one may think in terms of "active" leaders and "inactive" leaders at this level. The inactive first-level supervisors tended to be relations oriented (high LPC) and to have high esteem for their subordinates. This finding fits well with Fiedler's (1967, Chapter II) recent prediction that high LPC leaders who feel satisfied with their interpersonal relations with group members will sit back and take it easy; that is, they will be relatively inactive in dealing with the group. The remarkable thing in this case however is that the high esteem these first-level leaders felt for their subordinates was negatively reciprocated. Since the subordinate's satisfaction with the leader and the leader's esteem for the subordinate are two sides of the same relational interface one would expect them to be significantly positively related.

The present findings lend some confirmation to claims that different organizational levels will place different demands on leaders. While the authors do not expect the particular patterns discovered in the present study to generalize very far, they would expect differences in leadership demands to occur whenever the leadership situation differs sharply from level to level. In particular, two aspects of the present sample should suggest caution against generalizing these findings directly to other populations. First, one must consider that a nursing service may be a special kind of organization. Second, moving up through the levels of a hierarchy, n becomes smaller at each level. In this case, n at the second level dropped to 8. A correlation in such a small sample can be strongly influenced by one or two individuals. Therefore, skepticism is appropriate regarding reliability of the second-level analyses.

The present findings suggest several organizational management dilemmas and leave some unanswered questions. In selecting first-level supervising nurses, for instance, task-oriented (low LPC) applicants would seem to be preferable because in this situation they are generally active in leadership behavior. In selecting second-level supervisors, however, high LPC (relations-oriented) applicants may have a better chance of success. Most organizations as a matter of policy try to fill vacancies at upper levels of supervision with personnel who have demonstrated good performance at lower levels. In the present organization, this policy would lead to the promotion of low LPC first-level supervisors to second-level supervisory positions. If these nurses carry their high-activity leadership pattern up with them, they stand to alienate their new subordinates by showing too much Initiating Structure behavior. This problem could be compounded if only low LPC nurses are originally selected for supervisory positions at the first level. In that case, there would be no high LPC nurses available for promotion to second-level positions.

On the other hand, low LPC nurses, if promoted to second-level positions, might change their old pattern of showing high consideration and high structure in response to new situational conditions. Further research is of course needed to determine the extent to which upward mobile managers can perceive changes in their new situation as opposed to the old. What steps can be taken to ease this transition are still largely matters of conjecture.

REFERENCES

ARGYRIS, C. *Integrating the individual and the organization*. New York: Wiley, 1964.

BARNARD, C. I. *The functions of the executive*. Cambridge: Harvard University Press, 1950.

CARTER, L. F., HAYTHORN, W., SHRIVER, E., & LANZETTA, J. The behavior of leaders and other group members. *Journal of Abnormal and Social Psychology*, 1951, 46, 589–595.

DUBIN, R. Supervision and productivity: Empirical findings and theoretical considerations. In R. Dubin, G. C. Homans, F. C. Mann, & D. C. Miller (Eds.), *Leadership and productivity*. San Francisco: Chandler, 1965.

FIEDLER, F. E. A contingency model of leadership effectiveness. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. New York: Academic Press, 1964.

FIEDLER, F. E. *A theory of leadership effectiveness*. New York: McGraw-Hill, 1967.

FLANAGAN, J. C. Critical requirements: A new approach to employee evaluation. *Personnel Psychology*, 1949, **2**, 419–425.

FLEISHMAN, E. A. The measurement of leadership attitude in industry. *Journal of Applied Psychology*, 1953, **37**, 153–158.

GEORGOPOULOS, B. S., & MANN, F. G. *The community general hospital*. New York: Macmillan, 1962.

GIBB, C. A. Leadership. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 2. Cambridge, Mass.: Addison-Wesley, 1954.

HEMPHILL, J. K. Situational factors in leadership. Educational Research Monograph No. 32. Columbus: Ohio State University, 1949.

KATZ, D., & KAHN, R. L. *The social psychology of organizations*. New York: Wiley, 1966.

NILES, M. C. *Middle management: The job of the junior administrator*. New York: Harper, 1949.

PELLEGRIN, R. J., & COATES, C. H. Executives and supervisors: Contrasting definitions of career success. *Administrative Science Quarterly*, 1957, **1**, 506–517.

PFIFFNER, J. M., & SHERWOOD, F. P. *Administrative organization*. Englewood Cliffs, N. J.: Prentice-Hall, 1960.

PORTER, L. W., & HENRY, M. M. Job attitudes in management: V. Perceptions of the importance of certain personality traits as a function of job level. *Journal of Applied Psychology*, 1964, **48**, 31–36.

ROSEN, H. Managerial role interaction: A study of three managerial levels. *Journal of Applied Psychology*, 1961, **45**, 156–160.

SALES, S. M. Supervisory style and productivity: Review and theory. *Personnel Psychology*, 1966, **19**, 275–286.

SMITH, P. C., KENDALL, L. M., & HULIN, C. L. *Measurement of satisfaction in work and retirement*. New York: Rand-McNally, 1968.

STOGDILL, R. M. Personal factors associated with leadership. *Journal of Psychology*, 1948, **25**, 35–71.

STOGDILL, R. M., & COONS, A. E. Leadership behavior: Its description and measurement. Ohio Studies in Personnel, Research Monograph No. 88. Columbus: Bureau of Business Research, Ohio State University, 1957.

TERMAN, L. M. A preliminary study in the psychology and pedagogy of leadership. *The Pedagogical Seminary*, 1904, **11**, 411–451.

TRIANDIS, H. C. Categories of thought of managers, clerks, and workers about jobs and people in an industry. *Journal of Applied Psychology*, 1959, **43**, 338–344. (a)

TRIANDIS, H. C. Differential perception of certain jobs and people by managers, clerks, and workers in industry. *Journal of Applied Psychology*, 1959, **43**, 221–225. (b)

WALKER, H. M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

WONDERLIC, E. F. *Wonderlic Personnel Test, Form IV*. Northfield, Ill.: E. F. Wonderlic, 1959.

(Received December 18, 1967)

(Continued from page 393)

Brand Awareness: Differential Roles of Fittingness and Meaningfulness of Brand Names: Rabindra N. Kanungo*: Department of Psychology, Dalhousie University, Halifax, Nova Scotia.

Interaction of Achievement Cues and Facilitating Anxiety in the Achievement of Women: Wilbert J. McKeachie*: Department of Psychology, The University of Michigan, Ann Arbor, Michigan 48104.

Relation between Birth Order and Being a Beautician: Philip S. Very* and Joseph A. Zannini: Rhode Island College, 600 Mount Pleasant Avenue, Providence, Rhode Island 02908.

Psychological Concomitants and Determinants of Vocational Choice: Kenneth M. Kunert*: Department of Psychology, University of Detroit, Detroit, Michigan 48221.

Self-Esteem as a Moderator in Vocational Choice: Replications and Extensions: Abraham K. Korman*: Department of Psychology, New York University, 21 Washington Place, New York, New York 10003.

Relationship between Interview Structure and Inter-Interviewer Reliability in an Employment Situation: Donald P. Schwab* and Herbert G. Heneman, III: Graduate School of Business, The University of Wisconsin, 1155 Observatory Drive, Madison, Wisconsin 53706.

Automated Instruction for Vigilance Training: Dennis A. Attwood and Earl L. Wiener*: 708 Balra Drive, El Cerrito, California 94530.

---

* Asterisk indicates author for whom the address is supplied.

journal of applied psychology

Vol. 52, No. 6, Part 1 December 1968

RELATIONSHIP OF MEMORY OPTIMISM TO WORK COMPETENCY AND PERSONALITY VARIABLES

H. MELTZER and D. LUDWIG[1]

Washington University

The purpose of this study was to investigate the relationship of memory dynamics and work competency to personality variables and personified values. The source material used was the life and work stories of 143 workers which included their reports on outstanding pleasant and unpleasant experiences, responses to questions concerning personified values, and ratings on their competency as workers. In general, statistically significant differences were found in favor of memory optimists also being rated as more competent on the job, having a higher level of autonomy, and rating higher in interpersonal competence. Workers who had high economic values were less often memory optimists, less competent on the job, and tended to be older.

The purpose of this study was to investigate two problems related to personality variables—memory dynamics and work competency. In the area of memory dynamics the field is relatively barren as far as workers as *Ss* are concerned. Meltzer and Ludwig (1967) found that, although younger workers are significantly more often memory optimists than older workers, each age group expressed significantly more pleasant memories than unpleasant. These findings were interpreted to suggest that personality factors rather than the age factor as such are influencing memory dynamics involved. In the only study previously published on college students, Purcell (1952) found a relationship between personality factors and security feelings as measured by the Maslow Security-Insecurity Test and discovered that features of childhood and adult memories alike were also shown to be significantly related to security scores. This he interpreted as an initial advance toward detailing personality and memory functions. An earlier investigator, Waldfogel (1948), on the basis of a lengthy study of the affective characteristic of childhood memories, came to the conclusion that "more work should be done in the dynamics of memory optimism and memory pessimism and on personality differences between memory optimists and memory pessimists [p. 31]." One of the chief purposes of the present study was to try to satisfy the implied research need indicated in these studies by investigating the relationship of memory dynamics to personality variables in workers.

Another important phase of this study was to investigate the relationship of work competency to memory dynamics and personality variables. There is extensive literature on the relationship of work competency to job satisfaction and personality variables. Since the authors' concern in this phase of the study was to investigate work competency to personality variables, there will be no review of the literature on job satisfaction and performance. Instead, interested readers may refer to the oft-quoted review by Brayfield and Crockett (1955), Triandis's (1959) critique, and the comprehensive chapter devoted to this topic by Vroom (1964), particularly pages 181–187. One of the earlier studies concerned with personality variables rather than job

[1] An earlier version of this article was presented at the meeting of the Midwestern Psychological Association, Chicago, May 1967.

© 1968 by the American Psychological Association, Inc.

satisfaction is that of Georgopoulos (1957) in which he reported that workers who perceive that effective performance is a path to their goals are more productive. The stronger the need, the more will this relationship hold. In a study of blue-collar workers, Friedlander (1966) found no difference between high and low performance. He found that low performers found recognition through advancement to be more important and that high performers attached more importance to opportunities for self-actualization in white-collar government workers, but he found no significant differences in blue-collar workers. Guion and Gottier (1965) found very little indication of a significant relationship between personality test results and prediction of performance. In this present study an attempt was made to investigate work competency as rated by interviewers with two personality variables, autonomy and interpersonal competence as well as memory dynamics. Also explored was the relationship of personified values to memory optimism and work competency. In the light of the foregoing studies, as well as the authors' previous study, three hypotheses are made. These can be expressed as follows:

1. Memory optimists will generally be more adequate as workers, more autonomous, and will rate higher in interpersonal competence.
2. People who rated as more competent workers will tend to be memory optimists, more autonomous, rate higher in interpersonal competence, and typically will be younger.
3. Workers who are more economic minded as compared to those who are more political minded (in the Allport-Vernon-Lindzey sense) will less often be memory optimists, rate less competent on the job, and will tend to be older.

Since the nature of the study will be more clearly understood in the light of knowledge of the source material and *Ss* used, as well as the industrial setting of the community in which the study was made, next will be presented the method and procedures where this material is described.

METHOD

Subjects and Setting

The *Ss* used in this study were 143 workers in a paper converting industry located in a stable college community in upstate New York. This paper mill is the only plant of any size in the community and obtaining a job there at that time was considered by a good many of the people living in the community as the beginning of establishment. For the most part the workers were semiskilled and skilled, with large families and strong feelings about their positions with the company, jealousy over promotions of others, and pride in their own advance. The workers ranged in age from 19 to 78 and represented about one-third of all of the workers in the plant. The 143 were selected from the larger group because they were the only ones interviewed for whom data concerning memories of pleasant and unpleasant experiences as well as personified values were available.

The source material used was based on initial interviews with workers who were taken over by the new company. The interviewers were representatives of the new company and the attitudes toward them were extremely favorable because of dissatisfaction with the old company. The interviewers, in addition to the senior author, were graduates of Washington University with a minimum of 4 yr. of interviewing experience. The interview was structured in the following sequence (Meltzer, 1950):

1. Directed questioning about work, family, personal story, and present conditions or problems.
2. Recall of the most outstanding pleasant experience with question, "And now tell me what comes to your mind when you think of the most outstanding pleasant experience of your life."
3. Recall of other outstanding pleasant experiences.
4. Recall of the most outstanding unpleasant experience.
5. Recall of other outstanding unpleasant experiences.
6. Questions concerning personified values.
   (*a*) "Who is the greatest person that ever lived? Why?"
   (*b*) "Who is the greatest person living? Why?"
7. Questions concerning personified ideal or aspiration: "Of all the people you've ever seen, heard about or read about, whom would you rather be like? Why?"

Memory optimism was determined by predominance of pleasantness over unpleasantness in the experiences given by the worker, memory pessimism was determined by the predominance of unpleasantness over pleasantness, and indifferentism by an equal number of responses.

Work competency was based on the rating by the interviewers at the end of the interview. The interviewers had available for use the ratings of the respective supervisors. Scores for various ratings were as follows: 1 for D; 2, D+; 3, C; 4, C+; 5, B; 6, B+; 7, A; 8, A+.

Autonomy was determined on the basis of personal and work histories given by the worker in the interview, and a content analysis. The worker was given a plus (+) in autonomy if the histories contained statements indicating adequate security and

adjustment; minus (—) indicated insecurity and maladjustment. A zero (0) was given if no clearer assessment could be made. A second judge made independent ratings of the same factors with the resulting correlation of .6 between the two. (A 3 × 3 contingency table was used with the maximum correlation being .82.)

Interpersonal competency was also determined by a content analysis. A worker was given a + in interpersonal competence if the histories contained statements indicating that the worker got along well with others; if there were statements indicating a history of poor relations with others, a — was given; and a 0 was given if no clearer judgment could be made. The rating of the second judge correlated .64 in this measure. (A 3 × 3 contingency table was used with the maximum correlation being .82.)

Personified values were based on reactions to personifications of greatest ever lived, greatest living, and model.[2] The predominant value was assigned to each *S*, and of the seven values assigned only the four with sufficient frequency are reported. The ratings of the second judge correlated .68 for the two most frequent values.

RESULTS

Relationship of Memory Optimism to Autonomy and Interpersonal Competence

The workers were divided into two groups according to the memory optimists index; those with an index of +2 and above were put in the upper group (more optimistic). Those with an index of +1 and below were put in a lower group (less optimistic). The same procedure was followed in dividing the group for autonomy. Those who rated + were considered in the upper group here considered and those who rated — were put in the lower group. Those who scored 0 were not considered. (Since the middle group was therefore excluded in the calculations of the chi-square values for both the autonomy and interpersonal adequacy ratings, the total *N* is somewhat less than 143 in these cases.)

Table 1 gives a 2 × 2 comparison between the optimism and autonomy of the workers. The chi-square value of 10.97 ($p < .001$) and the corresponding phi coefficient of .35 indicate that those who have greater memory optimism also tend to be more autonomous or tend to be more secure and well adjusted.

[2] For details of procedure in the use of projective interviewing of which this is a phase, see Meltzer (1950).

TABLE 1

RELATIONSHIP OF MEMORY OPTIMISM TO PERSONALITY VARIABLES

Factor	Optimism group	Group: Upper factor *n*	Group: Lower factor *n*	ϕ	χ^2
Autonomy[a]	Upper	24	12	.35	10.97*
	Lower	18	39		
Interpersonal competency[b]	Upper	27	19	.02	.66
	Lower	34	31		

[a] $N = 93$.
[b] $N = 111$.
* $p < .001$.

The workers were divided into two groups according to interpersonal competency. Those who rated + were placed in the upper group and those who rated — were placed in the lower group. A 2 × 2 comparison of memory optimism and interpersonal competency is given in Table 1. The nonsignificant chi-square value of .66 and the corresponding phi coefficient of .02 must be interpreted as failing to support any relationship between memory optimism and interpersonal competence.

Relationship of Work Competency to Autonomy, Interpersonal Competency, and Age

To compare these variables the workers were divided into upper and lower groups for each factor. For the work competency variable, those with a C+ and above were put in the upper group, and those with C and below were put into the lower group. For the age variable, those above 40 were put into the upper group and those below 40 were put into the lower group. The divisions for the other factors are the same as given in Table 1. The comparison of all of these variables was made in a 2 × 2 fashion and the results are given in Table 2. Chi-square values were all significant and show that the more competent workers also tend to be more often memory optimists ($\phi = .19$, $p < .05$), tend to be more autonomous ($\phi = .40$, $p < .001$), tend to be more competent in interpersonal relations

($\phi = .23$, $p < .01$), and tend to be younger ($\phi = .20$, $p < .01$).

All in all, what the results reported in Table 2 indicate is that the more competent people rated higher in performance (work competence), also were more autonomous, rated higher in interpersonal competence, were younger, and were more likely to be memory optimists.

Memory Optimism and Personified Values

After the group of workers was rated according to dominant personified values in their lives, it was found that economic and political values classified nearly all of the workers. With one exception, then, the statistical treatment of the data employed these two divisions. The same divisions into upper and lower groups were made for the other variables as were given in the previous tables. The comparison of these variables was also made in a 2 × 2 fashion and the results are given in Table 3. This table shows that those workers who were classified as being more economic minded as compared to those who were more political minded were less often memory optimists (not significantly so), were rated as less competent on the job ($\phi = .19$, $p < .01$), and tended to be older ($\phi = .20$, $p < .01$). This table also shows that those few workers who were more religious minded as compared to those who showed more theoretical interests tended to be older ($p < .01$). The phi coefficient for this comparison is .84 and the small number of cases needs to be taken into account for interpretation of this figure.

TABLE 2

RELATIONSHIP OF WORK COMPETENCY TO PERSONALITY VARIABLES

Factor	Work competency group	Group: Upper factor n	Group: Lower factor n	ϕ	χ^2
Autonomy[a]	Upper	32	17	.40	14.50**
	Lower	11	33		
Interpersonal competency[b]	Upper	36	25	.23	5.83*
	Lower	18	32		
Age[c]	Upper	32	42	.20	6.04*
	Lower	44	25		
Memory optimism[c]	Upper	38	36	.19	5.15*
	Lower	20	49		

[a] $N = 93$.
[b] $N = 111$.
[c] $N = 143$.
* $p < .05$.
** $p < .001$.

TABLE 3

MEMORY OPTIMISM AND PERSONIFIED VALUES

Factor	Value	Group: Upper factor n	Group: Lower factor n	ϕ	χ^2
Optimism[a]	Economic	26	38		*ns*
	Political	25	32		
Work competency[a]	Economic	25	39	.19	4.28*
	Political	33	24		
Age[a]	Economic	40	23	.20	4.65*
	Political	25	33		
Age[b]	Theoretical	1	5	.84	8.57**
	Religious	6	0		

[a] $N = 121$.
[b] $N = 12$.
* $p < .05$.
** $p < .01$.

DISCUSSION

The purpose of this paper was to study the relationship between memory dynamics, personality, and performance variables in a field-study situation with a group of 143 workers as *S*s. The memory dynamic under consideration was the degree of memory optimism, the performance variable was the competency rating of the worker, and the personality variables included autonomy, interpersonal adequacy, and the value system of the workers. The age factor was also considered.

Looking first at memory optimism, a number of factors seem to be correlated with the degree of such optimism. The workers more optimistic about their past experiences also tend to be judged as more adequate on the job and more autonomous than do the less optimistic workers. It would seem from this that the success on the job for the workers

of this community is a very important factor in their memory optimism—or at least the two go together. The role of this factory in the community would seem to bear this out, for getting a job at this factory is something looked forward to by many in the area. The relation between autonomy and optimism is also a logical one and would seem to indicate a positive outlook on the part of these workers. In other words, those who feel more secure and adjusted can view their past positively and with less bitterness—thus showing memory optimism. This is in accord with Purcell's (1952) finding that relative frequencies of joy and fear memories are significantly related to security feelings.

Looking at the factors related to the rated job adequacy of the workers, some interesting relationships are also present. Those who are rated higher on job adequacy also tend to be more autonomous and interpersonally competent, tend to be younger, and tend to be more politically than economically oriented. Aside from actual ability, which was held constant in this study, there would seem to be a syndrome for the adequate worker. Good adjustment and the ability to get along well with others would certainly enhance the productivity of the worker, as this and many other studies bring out. As far as the younger worker's being more productive or more adequate, this would probably be a factor of the type of work required in this factory. The skills required to run the machines are not very complex, so there would not be an increase in job adequacy with more practice, whereas actual motor skills have been shown to decrease with age. The value system for the adequate worker is also an interesting finding. It would seem that the better worker is not working just for the money nor is he concerned primarily with increased pay (economic value), but is more interested in developing his skills and looking forward to advancement (political or power value). Such a person would also tend to look for ways of improving production and gaining prestige in the eyes of his superiors rather than concentrating on or complaining about his salary. This work attitude would tend to favorably influence the competency rating of the worker.

These findings remind one of statements made by Dobzhansky (1967) in an article in *Science* on "Changing Man":

> Optimists believe that ours is the best of all possible worlds. And pessimists are those who fear that the optimists are right. This is a flippant, but valid statement of truth. Optimism is often a result of ignorance of cold and unwelcome realities. There is, however, another kind of optimism, which is pessimism surmounted. The world is far from perfect, but it is not unalterable. I am tempted to call this evolutionary optimism [p. 409].

What he calls "evolutionary optimism," Meltzer and Ludwig (1967) have called realistic optimism. They reported responses of some people who gave no unpleasant experiences whatever, but whose life stories indicate them to be eternal or wishful optimists (who have repressed unpleasantness) rather than realistic optimists. Realistic optimists face unpleasantness as well as pleasantness as realities they react to and profit from. It is not surprising, therefore, to find a positive relationship between memory optimism and work competence as well as work competence and personality variables.

Looking specifically at the factors related to the value system of the workers, we have already seen that the political value is related to job adequacy. There are also significant differences in values according to the age of the workers. The older workers tend to be more economic and religious minded, whereas the younger workers tend to be more political and theoretical minded. Looking first at the older workers, it would seem that the threat of dependence that comes with the approach of retirement brings about a greater emphasis on the monetary value. Also the reality of death and the end of the existence in this world seems to bring about a greater emphasis on the religious values of life. The younger workers, with a life ahead of them, tend to look more for power and prestige, and with the threat of death more remote tend to concentrate on contemporary living rather than on religious values. Thus, in general, it would seem from this study that, by the use of a more clinical approach in the interviewing process, knowledge concerning the nature of memory optimists in relationship to performance ratings and personality variables and knowledge concerning the differen-

tial characteristics of more competent workers are obtainable. It must be added, however, that such a "critical approach" does add elements of subjectivity to the study.

REFERENCES

BRAYFIELD, A. H., & CROCKETT, W. H. Employee attitudes and employee performance. *Psychological Bulletin*, 1955, **52**, 396–424.

DOBZHANSKY, T. Changing man. *Science*, 1967, **155** (No. 3761), 409.

FRIEDLANDER, F. Motivation to work and organizational performance. *Journal of Applied Psychology*, 1966, **50**, 143–152.

GEORGOPOULOS, B. S., MAHONEY, G. M., & JONES, N. W. A path-goal approach to productivity. *Journal of Applied Psychology*, 1957, **41**, 345–353.

GUION, R. M., & GOTTIER, R. F. Validity of personality measure in personnel selection. *Personnel Psychology*, 1965, **18**, 135–164.

MELTZER, H. Memory dynamics, projective tests and projective interviewing. *Journal of Personality*, 1950, **19**, 48–63.

MELTZER, H., & LUDWIG, D. Age differences in memory optimism and pessimism in workers. *Journal of Genetic Psychology*, 1967, **110**, 17–30.

PURCELL, K. Memory and psychological security. *Journal of Abnormal and Social Psychology*, 1952 (Supple.), **47**(2), 433–440.

TRIANDIS, H. C. A critique and experimental design for the study of the relationship between productivity and job satisfaction. *Psychological Bulletin*, 1959, **56**, 309–312.

VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.

WALDFOGEL, S. The frequency and affective character of childhood memories. *Psychological Monographs*, 1948, **62**(4, Whole No. 291).

(Early publication received April 26, 1968)

Journal of Applied Psychology
1968, Vol. 52, No. 6, 429–431

RELATIONSHIP OF KUDER PREFERENCE RECORD VERIFICATION SCORES TO ADJUSTMENT:

IMPLICATIONS FOR VOCATIONAL DEVELOPMENT THEORY

FRANK A. NUGENT

Western Washington State College

High school males with inappropriately low Verification (V) scores on the Kuder Vocational Preference Record (KPR) were compared to high school males with acceptable V scores on the California Psychological Inventory (CPI) to investigate whether personal conflicts might be contributing to low V scores. Intellectual capacity and testing conditions were controlled. Incorrectly marked or incomplete inventories were eliminated. The final *N* was 256 boys with acceptable V scores and 83 with inappropriate ones. Students with acceptable V scores showed significantly higher mean scores on 15 of the 18 CPI scales. Thus, the V score might be considered 1 index of personal and social maturity. Assumptions that vocational choice is an integral part of total personality development and that vocational maturity relates to total maturity received some support.

The purpose of this investigation was to explore the relationship of the Verification (V) score of the Kuder Vocational Preference Record (KPR) to adjustment.

Kuder developed the V scale as a check on the confidence that can be given to a student's answers on the interest inventory. In the KPR manual he discusses possible reasons for V scores which fall outside the range (38–44) which he has conceived of as acceptable for meaningful interpretation of interest profiles (Kuder, 1957). Scores above 44 he points out are usually a result of not understanding directions since 44 is the highest possible score. The reasons given for scores below 38, however, are varied and less clear cut. The explanations for low deviance include not only the possibility of the student marking the answer sheet incorrectly, but also the possibilities of reading difficulty, disturbing test conditions, physical illness or emotional upset, careless or insincere responses, or an atypical interest pattern.

Students with V scores below the acceptable range, who were tested under adequate conditions, who were technically accurate in test taking, and who had sufficient intelligence to comprehend the vocabulary of the inventory, are the *S*s of this study. The focus of this investigation then is whether emotional factors are contributing to the inappropriately low V scores.

Support for the contention that a low V score is an indication of emotional conflict is limited and predominantly conjectural. Kuder's speculations about the relationship of the low V score to emotional factors have been echoed by some counselors operating on clinical hunches and observed behavior of counselors, but no studies relating personality characteristics or adjustment process to V scores of the Kuder profile are in the literature.

However, related assumptions that vocational choice is an implementation of the self-concept and an integral part of total personality development and that vocational maturity relates to total maturity have been important concepts of vocational development theory during the past decade (Super, 1957). If these assumptions are valid, then it is conceivable that high school boys who could not meaningfully respond to the choices of the KPR might be reflecting confusions of self-image due to distorted perceptions of themselves. Also, it is possible that conflicts with others significant in the environment might have led to inner tensions which rendered them less personally or socially mature than the average high school boy. In these contexts the V score of the Kuder might be considered one index of personal and social maturity.

Based upon Kuder's conjectures, the clini-

cal hunches of counselors, and the assumptions of vocational development theory, the following hypothesis was formulated: High school boys with V scores on the Kuder Vocational Preference Record within the acceptable range (V–AR) will show more favorable scores on a personality inventory than will high school boys with V scores not in the acceptable range (V–NAR).

METHOD

Kuder Vocational Preference Records (KPR) and California Psychological Inventories (CPI) were given in a 4-yr. high school to all males except those in special education classes. The town has a population of approximately 10,000 and is situated in the northwest region of the United States. The investigator supervised all the test taking to assure optimal testing conditions.

After eliminating individuals with IQs below 90, with incorrectly taken inventories, and with dissimulated CPI profiles (Gough, 1964), the final N was 256 boys with V–AR scores and 83 with V–NAR scores.

Mean scores on each of the 18 subscales of the CPI were obtained for both groups, and t tests for significance of mean differences were obtained.

RESULTS

In Table 1 the frequency distribution of these scores is shown. As can be seen from Table 2, high school boys with V–AR scores had significantly higher mean scores on 15 of the CPI subscales when compared to high school boys with V–NAR scores. The hypothesis was given considerable support.

Generally, on the basis of these significantly different mean scores, individuals with V–AR scores when compared to individuals

TABLE 1

FREQUENCY DISTRIBUTION OF V–NAR KPR SCORES

Score	f
16–17	2
18–19	0
20–21	1
22–23	2
24–25	4
26–27	5
28–29	3
30–31	9
32–33	10
34–35	15
36–37	32
Total	83

TABLE 2

COMPARISON OF SCORES ON THE CPI BETWEEN HIGH SCHOOL BOYS WITH V–AR KUDER SCORES AND THOSE WITH V–NAR KUDER SCORES

Scale	Kuder results	CPI scores	SD	t[a]
Measures of poise, ascendancy, and self-assurance				
Dominance	V–AR	45.42	12.45	−2.91**
	V–NAR	40.94	11.19	
Capacity for Status	V–AR	42.06	10.63	−3.39**
	V–NAR	37.41	11.48	
Sociability	V–AR	47.29	10.57	−4.07**
	V–NAR	41.94	9.83	
Social Presence	V–AR	51.78	10.74	−1.48
	V–NAR	49.76	11.00	
Self-Acceptance	V–AR	52.52	10.56	−1.83*
	V–NAR	50.16	9.20	
Sense of Well-Being	V–AR	40.69	13.28	−5.78**
	V–NAR	30.87	13.90	
Measures of socialization, maturity, and responsibility				
Responsibility	V–AR	43.78	10.49	−7.36**
	V–NAR	33.37	11.86	
Socialization	V–AR	50.94	10.49	−7.51**
	V–NAR	40.46	12.57	
Self-Control	V–AR	41.08	10.73	−3.02**
	V–NAR	36.96	10.81	
Tolerance	V–AR	41.73	11.48	−5.57**
	V–NAR	33.86	10.20	
Good Impression	V–AR	39.83	9.35	−2.07*
	V–NAR	37.33	10.03	
Communality	V–AR	50.96	11.32	−7.55**
	V–NAR	38.30	17.89	
Measures of achievement potential and intellectual efficiency				
Achievement via Conformance	V–AR	42.08	11.40	−6.53**
	V–NAR	32.60	11.71	
Achievement via Independence	V–AR	45.83	10.73	−3.95**
	V–NAR	40.54	10.10	
Intellectual Efficiency	V–AR	42.83	11.91	−7.55**
	V–NAR	31.05	13.55	
Measures of intellectual and interest modes				
Psychological-Mindedness	V–AR	48.82	9.90	−2.68**
	V–NAR	45.49	9.56	
Flexibility	V–AR	52.03	10.18	.54
	V–NAR	52.72	10.15	
Femininity	V–AR	48.17	9.22	−1.31
	V–NAR	46.70	7.71	

Note.—V–NAR = Verification scores not within acceptable range; N = 83; V–AR = Verification scores within acceptable range, N = 256.

[a] A one-tailed test was used since prediction was directional.

* $p < .05$.

** $p < .01$.

with V–NAR scores tend to be more self-confident, self-assured, and self-accepting (measures of poise, ascendancy, and self-assurance); more responsible and more emotionally and socially mature (measures of socialization, maturity, and responsibility); more adequately motivated, intellectually efficient, and resourceful (measures of achievement potential and intellectual efficiency); and more appropriately responsive to the demands and needs of others (Psychological-Mindedness).

DISCUSSION

Certain conjectures arise concerning students with V–NAR scores on the KPR when considered within the context of vocational development theory. Perhaps some of these students were subtly rebelling against the choice making demanded by the self-inventory as part of a general negative reaction or apathy to what they consider unreasonable or unattainable expectations from the environment. Possibly with some students an impasse in decision making occurred because of unresolved conflicts between their own inclinations and contradictory demands of parents and other authority figures. Such conflicts could contribute to self-doubts and confused self-perceptions with concomitant immobilization in making choices.

One further speculation is that some of these socially immature and personally ineffective students might not have had the opportunity to explore their interests or the world of work. In these situations, lack of experience or naïveté could be contributing to low V–NAR scores.

Counselors should be alert to the possibility that for some individuals low V–NAR Kuder scores may indicate pervasive personal problems of which difficulty of vocational interest choice is a part. If the broader areas of conflict are resolved, these individuals might then be able to evaluate their interests more realistically or might feel freer to systematically explore occupational fields consistent with their interests and abilities.

REFERENCES

GOUGH, H. *Manual for California Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1964.

KUDER, F. *Manual for Kuder Vocational Preference Record*. Chicago: Science Research Associates, 1957.

SUPER, D. *Psychology of careers*. New York: Harper, 1957.

(Received November 20, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 6, 432–439

MANAGERS' ATTITUDES TOWARD INTERACTION EPISODES

EDWARD E. LAWLER, III
Department of Administrative Sciences
Yale University

LYMAN W. PORTER
Graduate School of Administration
University of California, Irvine

AND ALLEN TENNENBAUM
University of California, Berkeley

An instrument was developed to measure managerial attitudes and behavior in communication-type interaction episodes. This self-recording form produced data on both managers' interaction activities during their working day and on how they felt about each episode that took place. The results showed that the managers studied were more favorable to self-initiated interactions than to other-initiated interactions. The data also showed that managers tended to evaluate the interactions they had with their superiors more positively than those that they had with their subordinates. The implications of these findings for communications within organizations were discussed, and the usefulness of the self-recording instrument for future research studies was considered.

There have been an increasing number of studies during the last 10 years that have focused upon managers and the managerial job (see, e.g., Vroom's 1965 review). Yet only a relatively small proportion of these investigations have studied the actual behavior of managers on their jobs. As Dubin (1962) has pointed out, "It is one of the curiosities of the literature on managerial behavior that we know almost nothing about it in any detail [p. 12]." The few studies that have been made of managerial behavior (e.g., Burns, 1954; Carlson, 1951; Davis, 1953; Dubin & Spray, 1964; Martin, 1959; Shartle, 1956) do provide some information about what managers actually do on their jobs, but they fail to enhance our knowledge of the psychological dimensions of such behavior. That is, they supply little or no data on why managers behave in a certain manner or on how managers react to different behavior episodes.

The main purpose of the present study was to explore the possibility of gathering data that will link certain kinds of managerial attitudes to specific behavioral episodes. In short, it was expected that if attitudes can be more closely linked to behavior then both the attitude data and the behavior data will become more meaningful.

Previous studies of managerial behavior have shown that managers spend a high proportion of their time talking to other people. Burns (1954) reports, for example, that managers spend 80% of all their time at work talking to others. This suggests that the study of the interaction episodes which take place during a manager's workday may be a fruitful place to begin studying some of the substantive issues concerned with managers' job attitudes. Furthermore, the study of these episodes also should contribute to an understanding of the communication process in organizations, particularly with respect to the questions of why certain communication patterns exist and why communication problems develop.

There are at least two important dimensions to consider in studying managers' reactions to interaction episodes. The first of these has to do with who initiates the interaction. Psychologically, there would seem to be a significant difference between a situation where a manager initiates an interaction and one where he has the interaction initiated for him by someone else. In the former situation the manager is likely to be in control of the situation and the interaction is likely to be held to satisfy his needs. In the latter case just the opposite may be expected to occur, and the manager is likely to be in a situation where he has less control and where the meeting is designed to fit someone else's needs. If, as is suggested by the research on small groups (Collins & Guetzkow, 1964), people generally feel more positively about those interactions that satisfy their needs and in

which they exercise control, then self-initiated interactions should be more positively evaluated than other-initiated interactions.

A second relevant dimension in determining a manager's reaction to an interaction episode concerns the position level of the other persons involved in the interaction. However, it is not immediately obvious as to whether an individual should typically feel most positively about an interaction with his subordinates or with his superiors. Arguments can be advanced for either prediction. When he talks to a subordinate, a manager is clearly in a higher status and higher power position, and this could put him in control of the situation and thereby create conditions that may lead him to feel positively about the interaction. A study by Hurwitz, Zander, and Hymovitch (1953), however, indicates that high-status and high-power individuals may tend to feel less positively about interactions than will low-status and low-power individuals. This study found that, after interactions, low-power individuals liked high-power individuals better than high-power individuals liked lower-power individuals. This suggests that superiors interacting with their subordinates might evaluate the episode less positively than would the subordinates, despite the fact that the former have more control in the situation.

In summary, the purpose of this study was both to develop and test an instrument for measuring managers' reactions to specific interaction behavior episodes, and to collect data that are intended to answer two substantive questions about managers' reactions to these interaction episodes: Does the initiator of an interaction feel more positively about the episode than does the recipient? Does the superior or the subordinate feel more positively about an interaction episode between the two?

METHOD

Research Instrument

Each respondent was asked to complete a number of self-recording forms. As can be seen from Figure 1, the form asks for factual information about a behavior episode as well as for the manager's attitudes toward the episode. The form itself is adapted from one used earlier by Burns (1954). It asks for some of the same factual information (some of which was not used in the present study) as the form developed by Burns, but it also includes attitude scales. This addition was crucial in terms of the present study since no previous studies have attempted to measure managers' attitudes immediately after an episode. This was of course done so that the attitudes could be directly related to specific episodes. The semantic-differential-type scales included were designed to tap the managers' global evaluation of the episodes.

The instructions for using the forms asked the manager to fill out a new form "at the conclusion of each behavioral episode that occurs during the working day." The term "behavioral episode" was defined as "any situation that has an integrity of its own" (e.g., a meeting, a telephone conversation, etc.). The manager was told that "at the end of the day, you should have a number of completed forms that, when considered together, will give an accurate picture of how you spent your time during the day." The instructions went on to explain each of the individual questions on the form. It was pointed out that in answering the first five questions it might be necessary to check more than one alternative. The managers were cautioned to check the words "superiors, subordinates, or peers" only when the interaction involved someone who was actually a member of their organization. Finally, for Item 6, the attitude item, respondents were instructed to "rate the episode on each of the five scales provided."

Sample

The sample included 105 middle and lower level managers from five organizations. Thirty-four of these managers were middle level and 71 of them were lower level managers.

The lower level managers were on the lowest level of management in their organizations and were generally the first line of supervision; in some cases individuals in this classification had few supervisory duties but were considered management individuals by the organization employing them. The middle level managers were in positions above the first level of supervision, but below the vice-presidential, company officer, or major departmental head level.

One of the five organizations studied was a large manufacturing plant with over 3,000 employees. Fifty of the managers studied came from this organization. The other 55 managers studied were employed by four local social service agencies. Included among the sample of social service managers were the directors and managers of a YMCA, a visiting nurse service, a girl scout council, and a city department of health. The 105 managers studied represented all the managers employed by the social service organizations and a random sample of about 40% of the lower and middle level managers employed by the manufacturing firm. A comparison between the demographic characteristics of the sample and the demographic characteristics of other samples of middle and lower level managers (e.g., Porter, 1961, 1962) revealed good comparability with but one exception. Thirty-one of the managers in

TIME
FROM_____TO_____

OTHERS INVOLVED_____

1. TYPE OF CONTACT

_____Letter or Memo
_____Discussion in Group
_____Discussion with one individual
_____Telephone
_____Records or Reports
_____Other

2. POSITION OF OTHERS INVOLVED

_____Superiors
_____Subordinates
_____Peers
_____Outsiders in Upper Management
_____Outsiders in Middle Management
_____Outsiders in Lower Management
_____Other Outsiders
_____None

3. TYPE OF ACTIVITY

_____Production/Operations
_____Sales/Distribution
_____Personnel/Industrial Relations
_____Public Relations/ Advertising
_____General Administration
_____Research and Development
_____Engineering
_____Finance/Accounting
_____Other

4. CONTACT INITIATED BY

_____Self
_____Other

5. PURPOSE OF EPISODE

	Information or Advice	Instructions or Decisions
Giving		
Receiving		
Creating		

6. ATTITUDE TOWARD EPISODE

Valuable	:_:_:_:_:_:_:_:	Worthless
Dissatisfying	:_:_:_:_:_:_:_:	Satisfying
Boring	:_:_:_:_:_:_:_:	Interesting
Precise	:_:_:_:_:_:_:_:	Vague
Challenging	:_:_:_:_:_:_:_:	Not Challenging

FIG. 1. Copy of the Self-Recording Form.

the present sample were females, while none of the other samples contained female managers. The female managers all worked for the social service organizations. Because of this and because of certain organizational differences between the manufacturing firm and the social service organizations, it was decided to analyze much of the data from these groups separately in order to determine if comparable findings would appear in two rather diverse populations.

Procedure

Meetings with the managers in each organization were held at the beginning of the workday on Monday. All 105 managers who attended the meetings agreed to participate in the study. A researcher first gave each participant a copy of the instructions for completing the self-recording form and a sample form. He then went over the instructions carefully with the managers and pointed out that the researchers were particularly interested in getting a form completed for each behavioral episode.

Questions were then invited from the managers with respect to the instructions given. Most of the questions were concerned with when it was appropriate to fill out a self-recording form. A number of specific incidents were discussed as examples. The re-

searcher continually stressed that he wanted to be sure that the completed forms would give a good picture of the interaction patterns of each manager during the week in question. (It had been arbitrarily decided in advance that, because of their frequency and highly specialized nature, managers' contacts with their secretaries would not normally be included.)

Each manager was then given 150 forms and five stamped envelopes addressed to the researchers at Yale University. The managers were told that they were to complete the self-recording forms for 1 wk. and that they were to begin as soon as the meeting ended. They were also told that at the end of *each day* they should put the completed forms in an envelope and mail them directly to the University. Thus, each manager was expected to mail five groups of completed forms to the University, one for each day of the study. This particular procedure was adopted for several reasons. First, it was hoped that this would emphasize to the managers the importance of completing the forms immediately after the episodes occurred. It also provided the researchers with a record of when the episode took place. Finally, it served to emphasize that this was not a company-sponsored project and that the company was not trying to exercise surveillance over their activities.

RESULTS

A total of 4,390 completed self-recording forms were obtained. The mean number completed (for 5 working days) by managers in the social service organization was 43.5, and the mean for managers in the manufacturing company was 39.9. Every manager completed at least 7 forms; the most that any manager completed was 112.[1] Table 1 shows the frequency of the different types of contacts in the two samples. As can be seen from the table, the managers in both samples reported a large proportion of their episodes involved verbal interaction with others. Only 11% of the episodes of the managers in the manufacturing company and 18% of the episodes of the social service organization managers did not involve verbal interaction.

TABLE 1

PERCENTAGE OF TYPES OF CONTACTS REPORTED

Type of contact	% manufacturing company	% social service
Groups	33	14
Individual	40	41
Telephone	14	28
Letter	6	8
Records	5	10

Of the total episodes reported for both groups, 74.9% involved contact with someone within the managers' own organizations. Table 2 shows the percentage of these within-own-organization contacts that were with superiors, peers, and subordinates. The managers reported that they generally tend to interact most frequently with their subordinates. This is not surprising since there are undoubtedly more subordinates for these managers to interact with than there are superiors. Burns' (1954) data also show this general tendency.

Table 3 shows the percentage of the contacts that are initiated by self, both, and others. The data show a relatively even balance between self and others, and they show that very few contacts are *seen* as being jointly initiated. Thus, in the analysis of the attitude effects of who initiates an interaction it was decided to focus on self- and other-initiated contacts. Additional analyses of the data (not shown in the table) indicated that the episodes reported were about equally spread out during the day from 8 A.M. to 4

TABLE 2

PERCENTAGE OF WITHIN-ORGANIZATION CONTACTS THAT INVOLVED SUPERIORS, PEERS, AND SUBORDINATES

Contact	% manufacturing company	% social service
Superiors	24.9	22.2
Peers	43.2	32.2
Subordinates	60.2	58.3

Note.—These percentages add up to more than 100 because more than one type of individual can be engaged in each interaction.

[1] The relatively wide range in the number of completed forms returned for the 5-day period raises the question of whether some respondents were neglecting to report some of their episodes. One way to check this, of course, would be to have observers on hand in the work situation to record the factual occurrence of interaction episodes. This was not feasible in the present investigation and ordinarily would not be feasible in most such investigations if any very large number of respondents were included in the sample. In the present study other available evidence—the fact that most managers' forms accounted for their total time during the working day—indicates that a high percentage of episodes were being reported.

TABLE 3

PERCENTAGE OF SELF-, OTHER-, AND BOTH-INITIATED CONTACTS

Initiator	% manufacturing company	% social service
Self	55	50
Both	3	3
Other	42	47

P.M. (except for the 12:00–1:00 lunch hour period).

The mean attitude scores toward all the episodes each manager reported were computed for him (Table 4). In this analysis the episodes were divided according to who initiated the episode and who was involved in it. Then the mean of each attitude scale was obtained for all the episodes reported by a manager. Finally, the scores for each of the managers were combined, and the mean attitude scores for all participants were obtained. Thus, the data presented in Table 4 consist of means which were computed from the individual manager's mean scores. Table 4 also reports a composite attitude toward each episode, computed by combining the data from the five attitude scales. Using the mean scores for each manager as the basic data in the study has the conservative effect of making N (105) the number of managers rather than the number of episodes reported. It also allows the responses of each manager to be weighted equally regardless of how many forms the manager completed.

TABLE 4

MEAN ATTITUDE SCORES FOR EACH OF THE FIVE SCALES: INITIATOR AND LEVEL OF PERSON CONTACTED

Attitude scale	Initiator	Contact		
		Superior	Peer	Subordinate
Valuable–worthless	Self	6.21	6.12	5.96
	Other	5.87	5.88	5.82
Satisfying–dissatisfying	Self	5.86	5.43	5.53
	Other	5.31	5.46	5.37
Interesting–boring	Self	5.48	5.47	5.29
	Other	5.38	5.41	5.18
Precise–vague	Self	5.76	5.53	5.69
	Other	5.37	5.25	5.43
Challenging–not challenging	Self	4.79	4.90	4.59
	Other	4.99	4.75	4.58
Composite for five scales	Self	5.59	5.52	5.43
	Other	5.46	5.38	5.27

Note.—Higher scores indicate more favorable responses. Standard deviations for the means reported in this table typically were close to a value of .90.

The data in Table 4 show, as expected, a general tendency for self-initiated contacts to be more highly evaluated than other-initiated contacts. This tendency was tested for statistical significance with a sign test (Table 5). For each manager a comparison was made between his average scores for his self-initiated and his other-initiated contacts. The results in Table 5 confirm the fact that in both samples there is a general tendency for the managers to feel more positively toward episodes they initiated than toward

TABLE 5

NUMBER OF MANAGERS WHO ARE MORE POSITIVE TOWARD SELF-INITIATED THAN TOWARD OTHER-INITIATED INTERACTIONS

Attitude scale	Manufacturing company				Social service organizations				Total			
	+	–	0	p	+	–	0	p	+	–	0	p
Valuable–worthless	36	11	3	.01	34	19	2	.10	70	30	5	.01
Satisfying–dissatisfying	31	17	2	.10	32	22	1		63	39	3	.01
Interesting–boring	27	19	4		28	24	3		55	43	7	
Precise–vague	29	16	5	.10	38	15	2	.01	67	31	7	.01
Challenging–not challenging	21	26	3		26	27	2		47	53	5	
Composite for five scales	34	15	1	.01	33	20	2	.10	67	35	3	.01

Note.—Plus sign indicates more favorable response to self-initiated episodes. Sign tests were two-tailed.

TABLE 6

NUMBER OF MANAGERS FOR WHOM THEIR ATTITUDE SCORES ARE MORE POSITIVE FOR INTERACTIONS WITH THEIR SUPERIOR THAN WITH THEIR SUBORDINATES

Attitude scale	Manufacturing company				Social service organizations				Total			
	+	−	0	p	+	−	0	p	+	−	0	p
Valuable–worthless	27	20	8		29	15	6	.05	56	35	14	.05
Satisfying–dissatisfying	28	20	7		29	14	7	.05	57	34	14	.05
Interesting–boring	33	16	6	.05	28	16	6	.10	61	32	12	.01
Precise–vague	29	18	8		21	24	5		50	42	13	
Challenging–not challenging	31	17	7	.10	27	17	6		58	34	13	.05
Composite for five scales	36	13	6	.01	27	20	3		63	33	9	.01

Note.—Plus sign indicates more favorable response to interactions with superiors. Sign tests were two-tailed.

episodes initiated for them by others. The only dimension that deviates slightly from this tendency is the challenging–not challenging one. However, for the total sample, the overall effect is highly significant as is shown by the results for the five-scale composite analysis.

The data in Table 4 also showed that there is a general tendency for managers to feel more positively toward contacts that they have with their superiors than toward contacts they have with their subordinates. The peers' contacts seem to fall between the superior and subordinate contacts. This tendency likewise was tested for significance with a sign test (Table 6). A comparison was made for each manager between his mean attitude toward all contacts he reported that involved his superiors and all those that involved his subordinates. As can be seen from Table 6, a significant majority of the managers reported more favorable attitudes toward the contacts that they had with their superiors than the contacts they had with their subordinates. Overall, this tendency appears to hold for both samples and for all of the attitudes studied.

If the findings that self-initiated communications and communications with superiors are evaluated highly are combined, this would suggest that the most positive attitudes managers hold should be toward self-initiated contacts with their superiors. On the other hand, the least positive should be toward other-initiated contacts with their subordinates. The data in Table 4 showed that this was the case. There was a significant difference between the mean composite score for self-initiated contacts with superiors ($\bar{x} = 5.59$) and the mean composite score for other-initiated contacts with subordinates ($\bar{x} = 5.27$) ($t = 3.2$, $p < .01$).

Table 7 presents the composite attitude scores (the mean for the five scales) for all the "dyadic interactions" that could be found in the data. Here, self-recording forms were included only where it was clear that forms had been obtained from both a superior and his subordinate who were reporting on a common episode. Thus, in these data, the reports of both people taking part in the *same* interaction could be compared. The previous data, of course, involved looking at the average reaction of managers to many different kinds of contacts (i.e., groups, dyads, etc.)

TABLE 7

MEAN ATTITUDE SCORES FOR DYADIC INTERACTIONS: INITIATOR AND LEVEL OF OTHER PERSON CONTACTED

Evaluator	Initiator		Grand M
	Self	Other	
Subordinate	5.78[a]	5.77[b]	5.78
Superior	5.71[b]	5.07[a]	5.42
Grand M	5.74	5.45	

Note.—Higher scores indicate more favorable attitudes.
[a] $n = 20$.
[b] $n = 24$.

while the dyadic data allow for the comparison of the attitudes of both participants to a specific interaction. As can be seen from Table 7, the initiator of the interaction feels more positively toward it than does the recipient, and the subordinate in the interaction feels more positively toward it than does the superior. Sign tests were computed between the attitude responses of each pair of participants and these trends proved to be statistically significant ($p < .05$). These results, therefore, serve to confirm the results reported above for the larger analysis which involved all interaction episodes by the total sample of managers.

DISCUSSION

Even though this study was essentially exploratory, the results have both substantive and methodological implications. Relative to the former, they provide information for analyzing communications in organizations and for explaining certain job-satisfaction patterns that appear in organizations. From a methodological point of view, they are suggestive of the fact that meaningful data about attitudes toward specific behavior episodes can be obtained by the use of a self-recording form of the type employed in this study.

The finding that managers tend to evaluate episodes they initiate more highly than episodes initiated for them was anticipated. It would seem to follow from three factors. First, the initiator of any contact usually has the most control in the situation. Second, the initiator normally does not institute a contact unless he feels that a worthwhile issue that is important to him needs to be discussed. Third, the initiator normally should feel more responsible for its taking place and might, therefore, be less likely to describe it negatively.

The finding that managers tend to evaluate contacts with their superiors more highly than contacts with their subordinates is understandable, even though it was specifically predicted. Part of the reason for this finding appears to be the fact that managers have relatively fewer contacts with their superiors compared to subordinates. Thus, a superior contact is relatively more "unusual" than a subordinate contact. In addition, the superior has reward power over his subordinates. Thus, a superior-subordinate interaction is likely to be a more significant event for a subordinate than for a superior. Apparently, as was suggested by the earlier study of Hurwitz et al. (1953), this factor is significant enough to overcome any tendency that superiors might have to evaluate the interaction more positively because of their superior control and power in the situation.

This finding also fits with some of the problems that are often discussed where superior-subordinate communication is involved. Haire (1956) and others have pointed out that, when talking with their superiors, subordinates often tend to overinterpret what is being said because they place too much importance on everything that is said. On the other hand, superiors may tend to do the opposite. The data from the present study suggest that this is in fact a real problem and that superiors do indeed tend to attach less significance to superior-subordinate interactions than do subordinates.

Given the results of the present study, it is not surprising that superiors often claim that they do not know what is going on in their organizations, because subordinates fail to communicate adequately. The results show that the least valued interactions are those where a superior evaluates an episode in which his subordinate comes to him for a discussion. If, as seems likely, the superior communicates any of these reactions to his subordinates, then it will not be long before the subordinate will "learn" to reduce his initiation of these interactions. In effect, the subordinate is not likely to be reinforced (sufficiently) for initiating communications and may very well decide that it is not worth coming to his boss with information, since his boss is not really interested.

Turning to methodological implications, the data seem to show that the self-recording form approach used in the present study is a useful way to measure managers' behavior and their reactions to the behavior. As evidenced by the large number of forms that the managers were willing to complete, and by the results obtained, it appears that managers did take the study seriously and completed them conscientiously. Many took the trouble to write

brief explanatory notes where they felt the categories on the form did not capture the essence of the contact. Some managers expressed the view that filling out the forms was an important learning experience for them, since the forms helped give an overview of the day's work. From a researcher's point of view, it was gratifying to find that the forms could be completed relatively quickly and easily and without seriously interrupting or distorting the day's typical work process. This approach also made it possible to measure the attitudes and behavior of a relatively large number of managers for at least a week's time. It would be difficult if not impossible to get comparable amounts of data by observation or any of the other commonly used approaches for gathering this type of behavioral data.

This instrument could also play an important role in developing what have been called "psychological" job descriptions. The typical job description provides a relatively poor behavioral description of the nature of a job. The self-recording form makes it possible not only to collect better descriptive data but also to get at some of the feeling and attitude dimensions of the jobs. It can pinpoint some of the psychologically difficult situations and interactions in the job, and, most importantly, the collection of such forms across a large number and wide range of interactions facilitates the linking of such interactions to organizational factors. Future studies might also look at the behavior and attitude patterns of managers who tend to be described as being effective and ineffective performers. In addition, actual behavior patterns might be related to the overall job satisfaction of the managers. Thus, a significant aspect of the present exploratory study is that it demonstrates the utility of the self-recording form and thereby suggests a number of future studies that can profitably be done in the realm of organizational psychology.

REFERENCES

BURNS, T. The directions of activity and communication in a departmental executive group: A quantitative study in a British engineering factory with a self-recording technique. *Human Relations*, 1954, 7, 73–97.

CARLSON, S. *Executive behavior*. Stockholm: Stromborgs, 1951.

COLLINS, B. E., & GUETZKOW, H. *A social psychology of group processes for decision-making*. New York: Wiley, 1964.

DAVIS, K. Management communication and the grapevine. *Harvard Business Review*, 1953, 31(5), 43–49.

DUBIN, R. Business behavior *behaviorally* viewed. In G. B. Strother (Ed.), *Social science approaches to business behavior*. Homewood, Ill.: Dorsey Press, 1962.

DUBIN, R., & SPRAY, S. L. Executive behavior and interaction. *Industrial Relations*, 1964, 3, 99–108.

HAIRE, M. *Psychology in management*. New York: McGraw-Hill, 1956.

HURWITZ, J. I., ZANDER, A. F., & HYMOVITCH, B. Some effects of power on the relations among group members. In D. Cartwright & A. Zander (Eds.), *Group dynamics*. Evanston, Ill.: Row, Peterson, 1953.

MARTIN, N. H. The levels of management and their mental demands. In W. L. Warner & N. H. Martin (Eds.), *Industrial man*. New York: Harper, 1959.

PORTER, L. W. A study of perceived need satisfaction in bottom and middle management jobs. *Journal of Applied Psychology*, 1961, 45, 1–10.

PORTER, L. W. Job attitudes in management: I. Perceived deficiencies in need fulfillment as a function of job level. *Journal of Applied Psychology*, 1962, 46, 375–384.

SHARTLE, C. L. *Executive performance and leadership*. Englewood Cliffs, N. J.: Prentice-Hall, 1956.

VROOM, V. H. *Motivation in management*. New York: American Foundation for Management Research, 1965.

(Received November 30, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 6, 440-446

EMOTIONAL MEANING OF FOUR TYPOGRAPHICAL VARIABLES[1]

ALBERT J. KASTL AND IRVIN L. CHILD

Yale University

The influence of 4 typeface variables (angular versus curved, bold versus light, simple versus ornate, serif versus sans-serif) on judgments of emotional meaning was studied in 40 college students. In addition to specific conclusions for each mood, it was found that, in general, moods such as sprightly, sparkling, dreamy, and soaring tend to be matched to curved, light, ornate, and perhaps sans-serif type; while moods such as sad, dignified, and dramatic are matched to angular, bold, and perhaps serif type.

Typefaces used in books or advertisements seem often to be chosen to emphasize or convey certain moods. This research was concerned with some of the physical attributes of type which apparently make mood communication possible.

The fact that certain typefaces can convey moods has long been established (Berliner, 1920; Burt, Cooper, & Martin, 1955; Davis & Smith, 1933; Ovink, 1938; Poffenberger & Franken, 1923; Schiller, 1935); among these earlier studies only those by Ovink and Davis and Smith make generalizations beyond individual typefaces. Ovink concluded that the light–bold dimension was primarily responsible for the emotional meaning of typefaces. Davis and Smith maintained that several additional factors were also responsible: size, condensation, slant (italic versus roman), and style of type (a complex group of factors distinguishing one family of type from another, as the difference between Caslon and Cheltenham type). Their pioneering experiment faced the complexity of the problem and represented a significant advance; yet it has several limitations:

(*a*) The categories employed are those traditionally used by typographers. It may be questioned whether these categories are sufficiently distinct to ensure that laymen respond to those features of type which determine traditional categorization.

(*b*) The results are of uncertain statistical significance.

(*c*) Only 13 typefaces were used, so that the categories are represented very poorly. (At the extreme, only one type represents the entire category of "expanded.")

A recent study by Tannenbaum, Jacobson, and Norris (1964) proceeds from criticism similar to ours of earlier work, but is primarily concerned with type characteristics different from those we consider. We learned of their work only after completing ours, and so could not design the present study in relation to theirs. Later, their findings will be compared with our own.

We chose for examination the following dimensions:

A. Angular versus Curved. For simple lines, an experiment long ago (Poffenberger & Barrows, 1924) supported the expectation that curves readily suggest quiet and gentleness; angles, harshness and seriousness. We expected a similar relationship with typefaces, both because of this experimental finding and because of the general use of this dimension in visual art.

B. Bold versus Light. This is the dimension stressed by Ovink (1938). Also, artistic usage suggests an important role for this variable.

C. Simple versus Ornate. This typeface dimension seems not to have been studied by psychologists, but its general role in art (cf.

[1] The authors are indebted to Stephen Dwoskin for technical advice on typography and to John D. Carroll for assistance in the statistical analysis. The investigation was carried out during the tenure by Albert J. Kastl of a National Science Foundation predoctoral fellowship. A more detailed statement of results, a classified list of the typefaces, and samples of their appearance have been deposited with the American Documentation Institute. Order Document No. 10004 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.25 for photocopies or $1.25 for microfilm and make checks payable to: Chief, Photoduplication Service, Library of Congress.

Barry, 1957 and Fischer, 1961) suggested its inclusion here.

D. Serif versus Sans-Serif. This dimension, too, has not yet been systematically studied for its possible emotional meaning. Serifs partly fill the gaps between characters and tie up the bulk of the word in a horizontal direction. An obvious hypothesis is that serifs will suggest forceful and stable moods.

The limitations indicated earlier were met by (*a*) having *Ss* themselves judge the physical attributes of the typefaces; (*b*) using appropriate techniques to evaluate the statistical significance of the effects; (*c*) employing a relatively large number of types incorporated in a factorial design. Thirty-two types were used. They fall in two sets, each set comprising variation in three of the four dimensions. Types 1–16 varied in Dimensions A, B, and C; Types 17–32 varied in Dimensions A, B, and D. In each instance the three dimensions were combined in an orthogonal design; each dimension had two levels, and each combination of conditions was represented by two different typefaces. Thus the same types constitute 32 samples of type for studying Dimensions A and B, and 16 each for studying Dimensions C and D.

METHOD

Subjects

The *Ss* were 40 students in introductory psychology, 20 men at Yale University and 11 men and 9 women at The City College, New York (CCNY).

Procedure

Preliminary testing was done to insure the physical validity of the dimensions. Eleven judges were asked to categorize each of about 40 types we had selected for possible use. Approximately 20 were classified as angular or curved, bold or light, and simple or ornate, and another 20 as angular or curved, bold or light, and serif or sans-serif. When disagreement appeared, the experimenters substituted new types until adequate agreement was reached (at least 9 of 11 judges concurring with our initial categorization). Thirty-two typefaces, 16 of each group of 20, were finally selected and made into slides. Each slide contained in a single typeface, and always in uppercase, the first 16 letters of the alphabet arranged in four rows and four columns. The letters were of fairly uniform height, but since shape varies between and within typefaces, width could not be standardized. The experimental procedure at Yale and CCNY was essentially the same. Twenty *Ss* were divided into four groups distinguished only by the order in which they judged various moods. To each group all 32 slides were shown eight times, once for judging each of eight moods. (Two random orders of presentation of the 32 slides were employed alternately, i.e., random order 1 was shown first, then random order 2, then 1 again, etc.) Each slide was shown for 8–10 sec., with a 3-min. rest period after the third and sixth showing of the entire set of slides. Group I judged the moods in the order a, c, e, g, b, d, f, h; Group II in the order c, e, g, a, d, f, h, b; Group III in the order e, g, a, c, f, h, b, d; and Group IV in the order g, a, c, e, h, b, d, f. Each of the moods is defined by a set of related adjectives; the sets employed are those used by Hevner (1935) and are presented in Table 3. In the text, we will refer to each set of mood words by the word which appears first in that set, both in Table 3 and in the form used by our *Ss*. We always intend reference, however, to the entire set. Moods a, c, e, and g were always judged before b, d, f, and h; and order varied only within each of these groupings. This arrangement seemed feasible because the similarity of the moods follows around in a circle so that the two groupings are as a whole highly similar to each other. It does not appear to have affected the reliability of the data. For Moods a, c, e, and g there are 16 significant main effects and 11 significant interactions, and for b, d, f, and h there are 16 significant main effects and 13 significant interactions.

When *Ss* viewed a projected typeface, they would indicate on a scale from 0 to 5 the degree to which they felt it conveyed the mood represented by the adjective set under consideration. A rating of 5 signified that the mood was conveyed very strongly; 4, considerably; 3, moderately; 2, somewhat; 1, slightly; 0, not conveyed at all.

The only difference between procedures at Yale and CCNY was that at Yale the slides were shown to the four groups of *Ss* separately; whereas at CCNY they were shown to all *Ss* at the same time, the groups being intermingled in the classroom so that no group had a better view of the screen than another.

Finally, at both Yale and CCNY, the slides were presented again for about 15 sec. each. The *Ss* were now asked to categorize the physical features of the typefaces by selecting one member of each of the following pairs as descriptive of that particular type: Types 1–16 were evaluated on the dimensions angular–curved, bold–light, simple–ornate; 17–32 on the dimensions angular-curved, bold-light, serif–sans-serif. Fourteen of the Yale *Ss* and 12 of the CCNY *Ss* participated in this part of the experiment. (Due to a procedural error, judgments of the Yale group on the simple–ornate and serif–sans-serif dimensions were not obtained.)

RESULTS AND DISCUSSION

If *Ss* give different mood ratings to typefaces of various categories, are they responding to the stimulus features intended to pro-

vide the basis for discrimination? An affirmative answer received advance support from the fact that preliminary judges agreed well with the experimenters' initial categorization. More pertinent, however, would be direct evidence that *Ss* themselves are able to discriminate on the intended basis. It was for this purpose that at the end of the experiment the *Ss* were asked to classify the typefaces. The results, in relation to what we intended, are indicated in Table 1. For example, as may be seen in the top line of the table, for 7 of the 16 types classified in advance as bold, over 95% of the *Ss* participating in this part of the experiment also classified them as bold; for 3 of the types classified in advance as light, over 95% of the *Ss* classified them as light; and so on. The other lines indicate the number of typefaces for which each lesser degree of agreement was found. Though agreement was not perfect, *Ss* taken as a whole differentiated consistently between types intended to be bold and those intended to be light, and between types intended to be serif and those intended to be sans-serif. On the other two dimensions, some individual types are not perfectly discriminated from all those intended to contrast with them; yet for each dimension the two sets of typefaces taken as a whole are very clearly differentiated.

To illustrate the statistical analysis of the mood ratings we will use the judgments made of the first 16 types for Adjective Group a. These 16 typefaces varied in Dimensions A, B, and C. The effects of these dimensions on Adjective Group a were studied in a five-way analysis of variance (A × B × C × E × S). (E represents the college variable, Yale or CCNY; and S, individual *Ss*). In Table 2 each of the main effects and interactions was compared with a term measuring its variation among *Ss* in order to obtain an F ratio (e.g., $F = A/AS$). Thus for the data from the first 16 types, an F was obtained for the A effect (23.86, $p < .001$), for the B effect (24.08, $p < .001$), and for the C effect (50.06, $p < .001$). Similarly, for the second 16 types, an F was obtained for the A effect (25.03, $p < .001$), for the B effect (48.86, $p < .001$), and for the D effect (8.13, *ns*).

The large size of these effects deserves emphasis. On Adjective Group a (sprightly), the 8 bold types among the first 16 had an average rating of 1.63, and the light types 2.41. An average rating of 1.82 was assigned to the angular types and 2.22 to the curved. For simple types the average was 1.57 and for ornate types it was 2.47. These are large differences; though the scale went from 0 to 5, the mean of all *Ss*' ratings on Mood a varied among typefaces only from .5 to 3.3, with a standard deviation of .78.

TABLE 1

CORRESPONDENCE BETWEEN SUBJECTS' JUDGMENTS OF PHYSICAL QUALITIES OF TYPEFACES AND THE EXPERIMENTERS' PRIOR CLASSIFICATION

Percentage of *Ss* whose judgments agreed with *Es*' prior classification	No. typefaces yielding each percentage of agreement, for typefaces originally classified by *Es* as							
	Bold	Light	Angular	Curved	Simple	Ornate	Serif	Sans-Serif
100	7	3	3	2	5	2	6	4
90	3	8	8	6	2	3	2	3
80	3	2	3			1		
70	1	2		3				
60	1	1		1	1			
50	1			2				
40			1	2		1		1
30								
20			1			1		
10								
0								

Note.—N = 26 for each of the first 4 columns of data and 12 for each of the last 4. Percentages have been rounded to the nearest multiple of 10.

TABLE 2

F TESTS SHOWING THE INFLUENCE OF THE TYPEFACE VARIABLES (AND THEIR INTERACTIONS) ON THE MATCHING OF MOOD-ADJECTIVE SETS

Typeface variables	Mood-adjective sets							
	a	b	c	d	e	f	g	h
Single variables								
Angular-Curved (A)	*23.86*** *25.03***	*39.78*** *15.53***	*27.51*** *1.79*	*29.38*** *5.19*	*3.33* *4.81*	22.04** 7.53	7.84 *12.78**	*2.83* *27.38***
Bold–Light (B)	*24.08*** *48.86***	*14.00** *44.80***	*101.13*** *110.30***	*52.91*** *35.66***	19.01** 22.68**	15.06* 12.25*	15.32** 34.30**	*6.85* *12.65**
Simple–Ornate (C)	*50.06***	*95.04***	*62.46***	*13.51**	3.32	.96	1.12	*19.69***
Serif–Sans-Serif (D)	*8.13*	*9.66**	*18.46***	*3.15*	3.02	1.26	25.52**	*0.11*
Interactions								
A × B	18.59** 3.77	5.28 3.08	33.32** 1.20	14.66* .46	.25 5.10	2.26 3.57	1.19 3.48	13.86* .70
A × C	1.05	7.65	.23	.48	4.61	8.73*	3.48	9.04*
A × D	25.85**	25.65**	11.81*	.96	.19	17.34**	16.74**	7.41
B × C	28.61**	14.40*	18.12**	14.01*	12.85*	5.21	13.61*	5.85
B × D	1.97	16.01**	1.69	19.62**	3.59	38.17**	11.45*	.02
A × B × C	.46	1.90	.09	11.52*	.06	3.66	.47	1.06
A × B × D	30.55**	6.49	4.78	9.66*	.42	.64	.19	.87

Note.—When results indicate that mood rating tends to be higher for angular, bold, simple, or serif type, the *F* is in roman. When the mood rating tends to be higher for curved, light, ornate, or sans-serif type, the *F* is in italics.
* $p < .01$.
** $p < .001$.

Table 3 provides a verbal summary of the dimensions and interactions which influence communication of each mood. The dimensions are represented in each instance by that end of the dimension which makes for communication of the particular mood. For interactions, a dimension is represented by that end whose presence enhances the main effect. Only those dimensions are included for which one of two *F*s for the main effect was found significant at the .001 level (or both at the .01 level) or the single *F* significant at the .01 level. We will now supplement the table by summarizing another way, reviewing the main effects and the interactions of each of the typeface dimensions.

Curvilinearity clearly tends to convey the moods expressed by the five Adjective Sets a, b, c, d, and h (sprightly, sparkling, dreamy, calm, and soaring). Angularity, on the other hand, conveys the mood of Adjective Set f (dignified). Thus there is considerable agreement between the Poffenberger and Barrows (1924) study on the feeling tone of lines and our findings about the effect of angularity and curvature in typography.

Light typefaces tended to be rated high by our *S*s on the moods represented by Adjective Sets a, b, c, and d (sprightly, sparkling, dreamy, and calm). Bold typefaces were rated high on the moods expressed by Adjective Sets e, f, and g (sad, dignified, and dramatic). These findings seem to agree well with the Ovink and Davis and Smith studies to the extent that we can properly match our moods with their somewhat different concepts. Ovink

TABLE 3

TYPEFACE CHARACTERISTICS SIGNIFICANTLY MATCHED TO SETS OF MOOD-ADJECTIVES, AND SIGNIFICANT INTERACTIONS

Set of mood-adjectives	Characteristics entering into	
	Main effects	Interactions
a[a]—sprightly, cheerful, bright, vivacious, happy, gay, joyous	1. curved 2. light 3. ornate	(light,* sans-serif*) (simple, curved) (bold)
b—sparkling, playful, merry, jovial, humorous, whimsical, fanciful, quaint	1. curved 2. light 3. ornate 4. sans-serif	(sans-serif*) (simple,* sans-serif) (bold) (curved,** light)
c—dreamy, dainty, delicate, light, graceful, lyrical, poetic, leisurely, gentle	1. curved 2. light 3. ornate 4. sans-serif	(light*) (simple, curved) (bold) (curved)
d[b]—calm, serene, soothing, tender, longing, romantic, plaintive, sentimental	1. curved 2. light 3. ornate	(light*) (simple, curved, serif) (bold*)
e—sad, pathetic, mournful, melancholy, depressing, gloomy, heavy, tragic	1. bold	(simple*)
f—dignified, spiritual, solemn, sober, serious, grave, earnest, lofty, philosophical	1. angular 2. bold	(simple) (serif*)
g—dramatic, forceful, vigorous, martial, ponderous, emphatic, majestic, exalting	1. bold 2. serif	(simple,* serif) (curved,* bold*)
h—soaring, triumphant, elated, exciting, impetuous, restless, stirring, spirited	1. curved 2. ornate	(angular)

Note.—When no asterisk appears in the Interactions column, the significant main effect (e.g., on the third line of the table, ornate) is separately significant at each end of the interacting variable (i.e., bold and light) at least at the .05 level, though more marked at the end named (i.e., bold).

[a] The triple interaction for Mood Group a suggests that for serif type the effect of boldness is enhanced by angular type and for sans-serif type the effect of boldness is enhanced by curved type. However, for both serif and sans-serif type the effect of lightness is enhanced by curved type.

[b] The first triple interaction for Mood Group d suggests that the effect of brightness is enhanced in curved types for both simple and ornate types. For ornate type the effect of boldness is enhanced in curved types; for simple types there is a slight but quite insignificant trend suggesting that the effect of brightness is enhanced in curved type, for serif type only. The second triple interaction suggests that for serif type the effect of boldness is enhanced by angular type; for sans-serif type the effect of boldness is enhanced by curved type.

* The main effect remains separately significant only for the end of the interacting variable named in the Interactions column; with the opposite level of the interacting variable the main effect of the first variable is not significant at even the .05 level, though it is still in the same direction.

** The significant effect of sans-serif type is maintained only for curved type. For angular type the effect is significantly reversed ($p < .05$).

noted that (1) the qualities of luxury and refinement are related to low pressure (light type), (2) soberness and economy to rather high pressure (bold type), and (3) force and strength to high pressure (bold type). This relationship between adjectives and type corresponds to our Adjective Sets c, f, and g, on which we found light (c), bold (f), and bold (g) types, respectively, to be rated high. Davis and Smith noted that the qualities of delicateness and beauty were related to nonbold (medium) type, while strength was related to bold type. This relationship corresponds to our Adjective Sets c and g, on which our *S*s gave high ratings to light and bold types, respectively.

Examination of the interaction of this dimension with the others suggests that (*a*) significant main effects of boldness are significant only if the type is simple rather than ornate, and (*b*) significant main effects of lightness are generally enhanced if the type is simple rather than ornate. The fact that the main effects of both lightness and boldness are greater in simple than in ornate type suggests that the present results have some generality for standard typography, since simple typefaces are usually employed.

Considering the simple–ornate dimension, we find that ornateness is judged to convey the moods of Adjective Sets a, b, c, d, and h (sprightly, sparkling, dreamy, calm, and soaring). There are no significant main effects in which simple type is rated higher on expression of a mood. Examination of the interactions of this dimension with the others shows that although simple type shows no significant main effects, it does enhance several main effects of other dimensions; ornateness never does this. This interaction indicates—to state more generally a point anticipated in the previous paragraph—that while simplicity by itself does not suggest a particular mood, it can render more conspicuous other distinctive features which do. Ornateness, on the other hand, apparently is so striking and distinctive in its emotional meaning that it diminishes the effects of all other typeface features simultaneously present.

Studying the serif–sans-serif dimension, we see that sans-serif type tends most strongly to suggest the moods of Adjective Sets b and c (sparkling and dreamy), and serif type the mood of Adjective Set g (dramatic). Although there are only these three significant main effects for this dimension, there are seven cases of significant interaction with main effects of other dimensions. For Adjective Sets a, b, and c (sprightly, sparkling, dreamy),

sans-serif type either enhances other main effects or is a significant main effect; for Sets d, f, and g (calm, dignified, dramatic), serif type either enhances the main effects or is itself a significant main effect. Comparing these two groupings of moods, we see that the first suggests freedom and high spirits; the second, restraint and seriousness. Apparently the serifs, by providing a horizontal linkage, facilitate communication of stability and emphasis as was originally hypothesized, and their absence has an opposite effect.

The main effects we have reported are not greatly limited by the interactions. In only two instances is a significant main effect of one variable reversed for one level of a second variable: For Adjective Group b (sparkling), the significant main effect of sans-serif is actually found only if the type is curved; if it is angular, serif types are instead rated higher on this mood, and this tendency is itself statistically significant ($p < .05$). For Adjective Group g (dramatic) the significant main effect of curved type obtained with the second 16 typefaces is actually found only if the type has serifs; here a slight reversal not approaching significance appears if the type is sans-serif.

When we look beyond these two exceptions, we find that the interactions do not cast doubt on the general validity of the main effects. For the 12 interactions marked with an asterisk in Table 3, the main effect of one variable is separately significant ($p < .05$) at only one level of the second variable, yet at the other level the effect still runs in the same direction. For the 16 interactions without asterisks in Table 3, the significant effect of the first variable remains separately significant ($p < .05$) for each level of the second, though the interaction shows the effect to be significantly larger for one level than for the other. Thus in a total of 28 cases of significant interaction the direction of the main effect is maintained for both levels of the second variable and in 2 it is not—evidence that the main effects are generally consistent in direction.

There were three significant triple interactions; only one of them (in Adjective Set d) affected the findings presented in Table 3. (This interaction indicates that only for serif types is the effect of lightness enhanced by curved types.) The fact that 13 of 16 triple interactions were insignificant suggests that the main effects and first-order interactions are on the whole maintained from level to level of the third variable.

Interactions of the college variable (Yale or CCNY) with the main effects and interactions presented in Table 3 were also computed. There were 88 such interactions; only 4 of them are significant at the 5% level, and of these only 1 suggests a reversal of a significant effect (it does so by so slight a margin as not to justify attention in the context of general agreement). The general agreement between colleges indicates that the findings are generalizable to other similar groups.

Some of these diverse findings can be organized into a single general scheme. Adjective Sets a, b, and c (and often d and h) form a group of moods which our *S*s tend to find well expressed by light, ornate, and possibly sans-serif typefaces. On the other hand, Adjective Sets f and g (and often e) form another group on which instead the angular, bold, and possibly the serifed types are rated high. The a, b, and c grouping consists of those adjectives which express happiness, joviality, delicacy; the f and g grouping expresses dramatic, forceful, serious, and philosophical moods. Each grouping consists of neighboring sets on Hevner's mood circle, in which adjacency signifies similarity. With fair adequacy, the two groupings may be characterized as suggesting femininity versus masculinity in connotative meaning. Viewed in this way, curvilinearity, lightness, ornateness, and absence of serifs convey femininity, while their opposites convey masculinity. This is certainly a helpful generalization for keeping the overall tenor of our results in mind, and the suggestion that sexual meaning is involved may be of considerable importance. (In another context, Jones and Lepson, 1967, have found black to be associated with masculinity, gray with femininity.) A masculine–feminine polarity provides, however, only a most general conception of results that are really much more complex. For example, Adjective Set h (soaring) seems on the whole to have more masculine than feminine connotation, and yet our *S*s rate high on it typefaces with characteristics we have

just designated as feminine (lightness, curvilinearity, and ornateness).

The relationships we find are probably brought about by many processes. As our study provides no basis for evaluating the relative importance of one and another process, we will not present here our speculations.

Our results may be put together with those of Tannenbaum et al. (1964) to support their conclusion that "there is a substantial basis for the use of typography as a code for the communication of connotative variations [p. 72]." Specifically, Tannenbaum et al. provide evidence that connotative meaning is influenced by slant (italic versus roman), by case (lower versus upper), and by type family. The serif–sans-serif dimension entered their research post hoc, and apparently they could not confidently ascribe any influence to it. Our research supplements theirs, then, by demonstrating the influence of this dimension. In addition, we show the great influence of three dimensions with which they were not concerned: angular–curved, bold–light, and simple–ornate. Tannenbaum et al., on the other hand, make another special contribution through comparing judgments by amateurs (comparable to our *Ss*) with judgments by professional typographers and by students of typography. In showing general agreement among these three kinds of judges, they demonstrate quite directly that the connotative meaning of typefaces can be utilized in communication of intended meaning from artist to public.

REFERENCES

BARRY, H. Relationships between child training and the pictorial arts. *Journal of Abnormal and Social Psychology,* 1957, **54**, 380–383.

BERLINER, A. Atmosphärenwert von Drucktypen. *Zeitschrift für angewandte Psychologie,* 1920, **17**, 165–172.

BURT, C., COOPER, W. F., & MARTIN, J. L. A psychological study of typology. *British Journal of Statistical Psychology,* 1955, **8**, 29–58.

DAVIS, R. C., & SMITH, H. J. Determinants of feeling tones in type faces. *Journal of Applied Psychology,* 1933, **17**, 742–764.

FISCHER, J. L. Art styles as cultural cognitive maps. *American Anthropologist,* 1961, **63**, 79–93.

HEVNER, K. Experimental studies of the affective value of color and lines. *Journal of Applied Psychology,* 1935, **19**, 385–398.

JONES, A., & LEPSON, D. S. Mediated and primary stimulus-generalization bases of sexual symbolism. *Journal of Consulting Psychology,* 1967, **31**, 79–82.

OVINK, G. *Legibility, atmosphere-value and forms of printing types.* Leiden: A. W. Sijthoff, 1938.

POFFENBERGER, A. T., & BARROWS, B. E. The feeling value of lines. *Journal of Applied Psychology,* 1924, **8**, 187–205.

POFFENBERGER, A. T., & FRANKEN, R. B. A study of the appropriateness of type faces. *Journal of Applied Psychology,* 1923, **7**, 312–329.

SCHILLER, G. An experimental study of the appropriateness of color and type in advertising. *Journal of Applied Psychology,* 1935, **19**, 652–664.

TANNENBAUM, P. H., JACOBSON, H. K., & NORRIS, E. L. An experimental investigation of typeface connotations. *Journalism Quarterly,* 1964, **41**, 65–73.

(Received December 11, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 6, 447–453

MANAGERIAL ATTITUDES TOWARD LABOR UNIONS IN A SOUTHERN CITY[1]

M. ALSIKAFI
University of Alabama

WALFRID J. JOKINEN
Louisiana State University

S. LEE SPRAY
Committee on Human Development, University of Chicago

AND

GEORGE S. TRACY
Louisiana State University

This paper deals with the development of a Guttman-type scale of attitudes of southern industrial managers toward organized labor and the extent to which the results of such quantified statements differ from those of industrial managers outside the South. A 7-item scale was obtained. A comparison of the outcomes of these items with measures of attitudes of industrial managers outside the South indicates the presence of a number of similarities and differences in patterns of responses related to matters of the role of the union, the problem of union shop, collective bargaining, and the process of decision making.

Writers in the field of industrial relations have long been interested in studying the attitudes held by managerial personnel toward organized labor. Attempts to develop quantified statements of these attitudes have been launched with varying degrees of success (Chalmers et al., 1953–54; Derber et al., 1960; Stagner, Chalmers, & Derber, 1958). Most of these attempts have been made in large industrial corporations in highly industrialized regions and the measures devised reflected this situation. No attempts have been made, on the other hand, to develop quantified measurements of managerial attitudes in southern industrial centers.[2] As a first step in the direction of accumulating evidence on this region, this paper presents the findings drawn from a study of managerial attitudes toward organized labor in one southern industrial center.

The experience of the industrial South with labor does not conform to many of the findings of industrial relations in other regions (Heberle, 1948; Simpson & Norsworthy, 1965). At least three factors are responsible for this: First, industrialization in the South began later than in other areas of the nation with significant consequences for management-labor relations; second, the social structure of the South, especially patterns of social stratification and the racial composition of the population strongly influenced the labor movement in the region (Heberle, 1959; Vance & Demerath, 1954); and third, the types of industries "creeping" southward produced few craft, white-collar, and professional jobs. The result has been an overrepresentation of the lower occupational strata (Simpson & Norsworthy, 1965).

To these factors one can add the fact that organized labor is weaker in the South than in other regions of the country. Both the proportion of organized labor and the rates of unionization of southern workers are below those of the nation as a whole. In 1953 the proportion of organized labor in the South was 17.1% of the total nonfarm labor force (compared to 32.6% for the nation). Until 1953 the average rate of unionization of the southern labor force was only 57% of the national rate. Although the South increased its proportion of unionized labor in recent years, it is still considered to be underorganized (Roy, 1965).

The effects of these characteristics upon labor-management relations have been the subject of a number of studies. Heberle (1948) has maintained that the spread of

[1] Revised version of a paper read at the annual meeting of the Southern Sociological Society, New Orleans, April 1966. The paper is based on data collected for a doctoral dissertation submitted by the senior author to the Department of Sociology, Louisiana State University, August 1965.

[2] The southern region used in this paper corresponds to the Southeast of Odum (1936) in his *Southern Regions of the United States.*

unionization in the South will result in industrial conflict different from that experienced in the North largely because of the traditionally paternalistic pattern of labor relations in the South. Confining his observations to southern textile industry, Gillman has noted that common customs and mutual interests so characteristic of this industry have created harmonious and sympathetic management-employee reliations which may be classified as "folksy" (Gillman, 1956). Roy (1965) analyzed changes in labor management in the South in terms of several factors that include the growth in influence of absentee-top management groups, the shift in the type of association between management and labor from "shirt-sleeved informality toward impersonal business relationships," and the replacement of early benevolent paternalism by pseudo-paternalism and coercion.

The foregoing studies clearly indicate that employee-management relationships in the South are markedly different from those of other regions but have recently undergone significant changes. Thus earlier forms of these relationships were characterized by folksiness, indigenous labor force, and a common culture. Benevolent paternalism provided these relationships with a strong shield against the encroachment of alien ideologies, thus preserving continuous harmony and cultural solidarity. More recent forms of management-labor relations have emerged during the past few decades with rapid industrial growth and the diversification of products. Basic to these new patterns is the growth of absentee ownership and "the replacement of Uncle Ben with college-trained officials" (Roy, 1965). Management in the South is now forced to deal with labor in terms of policies originating outside the region. Also, the recent industrial expansion has brought about a new surge of labor efforts to unionize workers of the region which has heightened labor's awareness of its self-interest and accelerated the breakdown of paternalism.

Differences in patterns of labor-management relationships between the South and the rest of the country are subtle and require a great deal of systematic analysis. In addition, southern industrial centers are currently experiencing a variety of types of relationships due to differences in level of industrial growth, type of product, and geographical location that it would be difficult to speak of an inclusive pattern. Preliminary attempts must be made, however, to reveal some aspects of these differences so that more comprehensive studies could be undertaken in the future. In this paper the field of attitudes of industrial managers toward organized labor is explored. The problem of the study is twofold: (*a*) to assess the kinds of attitudes industrial managers in the South hold toward organized labor, and (*b*) to assess the extent to which southern managerial attitudes differ from those held by managers in other regions. The strategy employed to accomplish these tasks was to develop a Guttman-type scale of attitudes of industrial managers in a southern city and then to compare the findings with available scales and other indexes of attitudes held by management outside the South. Since many of the instruments used in other studies contained items referring to local situations, they were not suitable for research in the South. For this reason, an attempt was made to develop a scale which would yield the desired data on southern managerial attitudes while at the same time being suitable for regional comparisons. Given the importance of this methodological problem, the technique used in scaling and the technical characteristics of the scale will be described in some detail.

The data were gathered from five bulk raw-material processing plants located in an administrative-industrial center in the Southern United States with a metropolitan area population of about 275,000. These corporations have some form of labor contract with either an independent labor union or a union affiliated with the AFL-CIO. Seventy lower-level officials and 20 top management executives employed by these corporations were interviewed. Although separate interview schedules for each of the two levels of management were prepared and used throughout the study, items making up the scale presented here were identical for the two managerial levels.

ATTITUDES TOWARD LABOR UNIONS

Items of the Scale

Attitudes of management toward labor unions were measured by seven items, each

containing four-response categories. These categories were assigned weights ranging from 0 for strongly favorable to 3 for strongly unfavorable. The items are as follows:

1. Labor unions not only better their members' economic and social conditions but also contribute to the welfare of society.
2. The drive to organize workers is a legitimate right for labor unions.
3. Collective bargaining is a necessary and important function of labor unions that should be preserved.
4. Labor unions are reasonable in their drive to recruit workers into their organizations.
5. Labor unions' demands regarding wages, hours of work, and working conditions, are in most cases, fair and reasonable.
6. Labor's demand for the union shop is a legitimate right for them.
7. Labor unions should have a share in making decisions in the corporation.

Scaling Technique

Using the scalogram board technique (Guttman, 1950), responses of the two groups of managers to the above items were scaled in one unit. All items were dichotomized in the second trial. The items formed a "true scale" since a coefficient of reproducibility of .90 was obtained, and the errors were randomly distributed (Guttman, 1950). The final order of the items was as follows: 1, 2, 3, 5, 6, 4, and 7.

The marginal frequencies of the seven items are shown in Table 1. Item 7 was retained although it failed to meet Guttman's criterion that an item should have 90 to 10% marginal frequencies to be included in a 10-item scale.

TABLE 1

MARGINAL FREQUENCIES OF THE ITEMS OF THE ATTITUDE SCALE OF UPPER AND LOWER MANAGEMENT OFFICIALS

Item[a]	Marginal frequencies percent	
	Disagree	Agree
1	18	82
2	23	77
3	52	48
5	58	42
6	61	39
4	72	28
7	93	7

[a] Items are listed according to their final order in the scale.

TABLE 2

DISTRIBUTION OF LOWER AND UPPER MANAGEMENT OFFICIALS IN THE ATTITUDES SCALE TYPES

Perfect scale type	Scale items[a]							Respondents	
	1	2	3	5	6	4	7	No.	%
I	x	x	x	x	x	x	x	11	12.22
II		x	x	x	x	x	x	12	13.33
III			x	x	x	x	x	22	24.44
IV				x	x	x	x	10	11.11
V					x	x	x	4	4.44
VI						x	x	5	5.55
VII							x	26	28.88
Total								90	100.00

[a] The symbol x designates disagreement with the respective item.

It was decided that marginal frequencies of 93 to 7% are close enough to the minimum requirement which is, in itself, an arbitrary limit.

Scores in the scale ranged from 0 to 7. Only one respondent scored 0 and one other respondent scored 7. Both respondents were lower managerial officials. Upper management officials contributed fewer than their pro rata errors (11% of the total number of errors in the scale). Their scores, however, were distributed in a consistent pattern.

Seven scale types were established with respondents ranging from 12% in the first scale type (classified as having the strongest negative attitudes toward labor unions) to 3% in the seventh type (considered to have the strongest prounion attitudes).[3] Table 2 shows the number of respondents in each scale type.

The theoretical framework of the scale includes the assumption that matters of agreement and disagreement between labor and management can be used to identify management feelings toward labor unions. Matters like the role of unionism in industrial society, the demands made by unions to improve workers' social and economic conditions, the significance of collective bargaining, labor's

[3] Stressing negative rather than positive attitudes toward organized labor is intentional since the paper is derived from a study of a set of hypotheses dealing with the impact of work orientations of managers upon their negative attitudes toward labor unions.

demand for the union shop, and labor participation in making decisions in the corporation were considered a valid basis for assessing attitudes of both upper and lower management personnel toward labor unions.

Polarization of attitudes and the likelihood of an attitudinal continuum is supported by the fact that the majority of the respondents agreed with the statement that labor unions have a vital role in the growth and welfare of modern society (Item 1), and disagreed with the statement that unions should have a share in making decisions in the corporation (Item 7). Between these two extremes, responses fell into a gradient pattern with the number of disagreements per item decreasing gradually toward Item 1 and increasing toward Item 7.

The final order of items in the scale reveals that respondents considered the existence of labor unions to be important to society. They also agreed that union efforts to recruit members are legitimate. On the issue of collective bargaining, however, responses leaned more to the negative side of the scale. Relatively high disagreement did not begin until respondents were asked to express their feelings toward the union shop. Disagreement became greatest (most widespread) when opinions were expressed about the legitimacy of labor to have a share in making decisions in the corporation.[4]

Judging respondents on the basis of their scores in the scale, four levels of attitudes were established: high negative, moderately high negative, moderately low negative, and low negative. The distribution of respondents in these categories is shown in Table 3. The table shows that the majority of respondents are concentrated in the two middle categories and that the two extremes have almost equal numbers of cases. The authors are disposed to conclude that southern managers have moderately negative attitudes toward organized labor.

The distribution of negative attitudes among managers of the five corporations is shown in Table 4. More than two-thirds of the managers in Corporation A show low-negative attitudes. On the other hand, Corporations B, C, D, and E are characterized by high-negative attitudes with varying degrees of intensity. The striking thing about this distribution is the relationship it bears to size of the corporation. Specifically, the largest corporation has the least negative views while the "medium-sized" and small corporations have the strongest negative views. While this finding may be an artifact of the specific union-management relationships in the various organizations, the authors were unable to find evidence to support this view. They are led to conclude, then, that factors associated with size of the organization influence managerial attitudes toward unions. Specifically, the less negative views found in the largest corporation may reflect the fact that labor-management relations have reached a highly stabilized pattern based upon common goals

TABLE 3

CLASSIFICATION OF MANAGEMENT BY ATTITUDES TOWARD LABOR UNIONS

Attitude	Score	No.	%
High negative	7,6	13	14.44
Moderately high negative	5,4	38	42.22
Moderately low negative	3,2	27	30.00
Low negative	1,0	12	13.33
Total		90	100.00

TABLE 4

DISTRIBUTION OF LEVELS OF NEGATIVE ATTITUDES IN THE FIVE CORPORATIONS

Corporation	Approx. no. employees	Attitudes toward labor unions				Total
		High negative		Low negative		
		No.	%	No.	%	
A	2200	6	26.09	17	73.91	23
B	720	18	72.00	7	28.00	25
C	700	12	66.66	6	33.33	18
D	125	9	56.25	7	43.75	16
E	100	6	75.00	2	25.00	8
Total		51		39		90

[4] Using Schuessler's Test I, it is concluded that the observed frequencies of the scale types exceed the frequencies expected by chance. See Schuessler (1961) for fuller details.

and mutual interests, and that leadership of both sides has developed specialized forms of communication to channel differences between them. It is also possible that unions constitute less of a threat to managers of large organizations than is true in smaller corporations. On the other hand, it seems likely that the "paternalistic" pattern of labor relations discussed by Heberle and the subsequent development of "pseudo paternalism" and "coercion" suggested by Roy, is more likely to be a strategy employed by the middle-sized and smaller corporations. If this is indeed the case, then the strong negative views found in the middle-sized and small corporations may reflect the managerial reaction to the strong conflict provoked by labor-management disagreements in such a system.

Attitudes and Organizational Commitment

Previous studies have examined the correlation between attitudes toward labor unions and liking or disliking for the company in order to demonstrate the existence of what has been called a "common attitudinal climate" among industrial workers and managers. Chalmers and his associates found that where industrial workers showed a liking for management, they expressed positive feelings toward unions and vice versa (Chalmers et al., 1953–54). On the other hand, tests of such relationships among industrial managers showed an absence of significant association (Stagner et al., 1958).

In the current study attempts were made to test the association between attitudes of *lower-level* managers toward labor unions and levels of commitment to the employing corporation. Organizational commitment was measured by responses to seven items dealing with feelings of managers regarding their work in the company, with their willingness to help the company during a crisis, and with their sense of identification with the corporation. The scalogram board technique was used to rank these items. In the final trial only five items were scalable with a coefficient of reproducibility of .90. Because the number of items was small and the requirements of a "true scale" were not fulfilled, details of the scaling procedures are not reported here.

The correlation between negative attitudes toward labor unions and organizational commitment indicates that the relationship is not significant.[5] Obviously this finding means that respondents who have high negative attitudes toward labor unions are not committed to their organization. While this may be correct, the authors must point out that their measure of organizational commitment may have influenced the outcome of the test. On the other hand, they take refuge in the finding of the Illini City study which found a "common attitudinal climate," and conclude that managers who have strong negative attitudes toward labor tend to lack strong commitment to the organization.

DISCUSSION

How do the attitudes of southern industrial managers toward labor unions compare with the views held by those outside the South? Here it is instructive to compare the findings of the results of two items of the scale with findings of a Gallup poll made public at almost the same time the data of the study were collected (Gallup, 1965). The poll asked a national sample: "In general do you approve or disapprove of labor unions?" In the Gallup sample 23% of the nonunion respondents disapproved of labor unions. In the current study of managers 18% disagreed with the statement: labor unions "not only better their members' economic and social conditions but also contribute to the welfare of society."

A second question in the Gallup poll asked: "Do you think a person should or should not be required to join a union if he works in a unionized factory or business?" Of the nonunion respondents in the poll, 56% replied negatively. The present sample of southern managers elicited 61% negative replies to Item 6: "Labor's demand for the union shop is a legitimate right for them."

It should be noted that the differences in percentages between southern managers and nonunion Gallup poll respondents for both items is rather small. In both instances the 95% confidence interval around the southern managerial sample proportions includes the proportion from the Gallup poll. Although

[5] $r = -.035$; $p(z = -.417) > .05$.

the specific wording of the items is somewhat different, in both cases the meanings are similar. Evidently southern managers do not disagree greatly with Gallup's nonunion respondents.

This should not be surprising, nor does it necessarily contradict Stouffer's findings that Southerners are less tolerant than the rest of the nation. Stouffer (1955) found that the amount of schooling and type of community (rural-urban) of his cross-section respondents influenced their level of tolerance.[6] Thus, the higher the educational level the more likely the individual is to show tolerance of nonconformists. Similarly, urban respondents tended to score relatively higher on the tolerance scale than rural respondents. In the current study, respondents are urban residents and have achieved higher levels of education than the "average" Southerner. Thus, the relatively high proportion of southern industrial managers approving of labor unions may reflect a tolerance that is due to their educational achievement and urban residency.[7]

Viewed from a different perspective, the relatively mild reaction of southern management to the existence of labor unions, especially in large corporations, can be attributed to the weakness of unionism in the South. It was pointed out earlier that southern labor unions have not reached the strength achieved by unions outside the southern region.[8]

The findings of the present study are most meaningful when compared to similar studies of managerial attitudes outside the South. The most comparable study is one in which Stagner et al. (1958) developed an 8-item Guttman scale to measure attitudes of managers to labor unions and a 9-item scale of union attitudes toward management. The authors found that in most organizations "management representatives were either moderately or very favorable toward the union" (Derber et al., 1960). They also found that managers were favorably disposed toward unions when asked to respond to more specific issues such as the role of the union in influencing the manner in which the organization was run, whether or not union claims were considered reasonable, and whether or not the amount of power enjoyed by union representatives was appropriate.

The findings of the Stagner et al. study differ from those reported in this study in several respects. First, in contrast to the generally favorable attitudes of the Northern managers, the Southern managers had a generally negative attitude toward organized labor. Second, even more pronounced differences exist between the two groups when the specific items of the two studies are found. While responses to the item dealing with contract negotiations "tended to be concentrated at the more favorable or friendly end of the scale" in the Stagner et al. study, responses of the Southern managers to the question dealing with collective bargaining clearly lean toward the negative side of the scale. (See Table 1.) To the question of labor unions' power and their interference in areas of production and discipline the general pattern of responses reported in the Stagner et al. study is favorable although "extensive management criticism was apparent in many establishments." The current study, on the other hand, shows a rather overwhelming majority of negative responses to the question dealing with the legitimacy of labor's demands for a union shop and "a share in making decisions in the corporation."[9]

SUMMARY AND CONCLUSIONS

In the absence of an instrument for measuring attitudes of southern managers toward unions, this paper has attempted to develop

[6] By tolerance Stouffer means "willingness to grant certain rights to people whose views might be disapproved, such as the right to speak, the right to hold certain kinds of jobs, etc." See Stouffer (1955) for details.

[7] The educational factor should probably be given more weight since urbanism by itself, as Stouffer points out, did not account for a significantly high level of tolerance in the southern region.

[8] The South is still considered by one union organizer of being the "new frontier" for organized labor. See Conney (1961).

[9] Scale items in the Stagner et al. study have a great deal of resemblance to the corresponding items of the current study. The basic difference is that questions in Stagner's study ask reaction of manager to relationships between themselves and the union which exist in actuality whereas items of the current study are simple declarative statements overly implying neither normative nor personal preference.

a Guttman-type scale for industrial managers in a southern city. The instrument developed met the requirements of a "true scale" as specified by Guttman. Perhaps one reason for the success of this scale is that the items presented to the respondents were generalized statements. This confirms the Stagner et al. conclusion that generalized items tend to scale better than specific items. The scalability of the items used in this study leads the authors to conclude, with the Stagner et al. study, that a Guttman scale represents a highly reliable way to ask managers "do you approve of the union?" (Stagner et al., 1958).

A comparison of the characteristics of the scale developed on southern managers with those developed on populations outside the South revealed a number of important differences. It is probably correct to conclude that most of the differences arose when questions focused on specific issues. The instrument itself seems to be more applicable to small corporations in areas not highly industrialized, and in regions where unionization has not dominated the modes of grouping employees.

A final word is in order. The findings of this study must be qualified by one limitation. That is, some of our "southern" managers, especially top executives, may not be southern in origin. It was pointed out that absentee ownership has become an important factor in shaping labor-management relations in the South. Given the fact that most of the corporations included in this study are absentee owned, and that the rate of geographical mobility of industrial executives is high, it is possible that some of our findings may reflect the influence of nonsouthern managers. In fact, the major cleavage in southern industrial organization may involve a difference between top and middle management with regard to their views on organized labor which is a function of the distinction between large absentee-owned corporations and smaller indigenous organizations.

REFERENCES

CHALMERS, W. E., ET AL. *Labor-management relations in Illini City.* (2 vols.) Champaign: Institute of Labor and Industrial Relations, University of Illinois, 1953–54.

CONNEY, R. B. The modern South: Organized labor's new frontier. *The American Federationalist,* 1961, **63**, 15–19.

DERBER, M., ET AL. *The local union-management relationship.* Urbana: Institute of Labor and Management Relations, University of Illinois, 1960.

GALLUP, G. Public deemed closely divided regarding right-to-work laws. *The Sunday Advocate* (Baton Rouge, La.), June 13, 1965. p. 17.

GILLMAN, G. *Human relations in industrial Southeast.* Chapel Hill: University of North Carolina Press, 1956.

GUTTMAN, L. The basis for scalogram analysis. In S. A. Stouffer et al. (Eds.), *Measurement and prediction.* Princeton: Princeton University Press, 1950.

HEBERLE, R. Social consequences of industrialization of southern cities. *Social Forces,* 1948, **27**, 30–34.

HEBERLE, R. The changing social stratification of the South. *Social Forces,* 1959, **38**, 42–50.

NOLAND, W. E. Technological change and the social order. In J. C. McKinney & E. Thompson (Eds.), *The South in continuity and change.* Durham, N. C.: Duke University Press, 1965.

ODUM, H. W. *Southern regions of the United States.* Chapel Hill: University of North Carolina Press, 1936.

PELLEGRIN, R. J., & COATES, C. Absentee-owned corporations and community power structure. *American Journal of Sociology,* 1956, **61**, 413–419.

ROY, D. F. Change and resistance to change in the southern labor movement. In J. C. McKinney & E. T. Thompson (Eds.), *The South in continuity and change.* Durham, N. C.: Duke University Press, 1965.

SCHUESSLER, K. F. A note on statistical significance of scalogram. *Sociometry,* 1961, **24**, 312–318.

SIMPSON, R. L., & NORSWORTHY, D. R. The changing occupational structure of the South. In J. C. McKinney & E. T. Thompson (Eds.), *The South in continuity and change.* Durham, N. C.: Duke University Press, 1965.

STAGNER, R., CHALMERS, W. E., & DERBER, M. Guttman-type scales for union and management attitudes toward each other. *Journal of Applied Psychology,* 1958, **42**, 293–300.

STOUFFER, S. H. *Communism, conformity and civil liberties.* Garden City, N. Y.: Doubleday, 1955.

VANCE, R. B., & DEMERATH, N. J. *The urban South.* Chapel Hill: University of North Carolina Press, 1954.

(Received December 15, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 6, 454–456

RELATIVE EFFECTIVENESS OF PROGRAMMED TEXT AND TEACHING MACHINE AS A FUNCTION OF MEASURED INTERESTS

D. D. CAHOON

Auburn University

LARS P. PETERSON AND CHARLES G. WATSON [1]

Veterans Administration Hospital, St. Cloud, Minnesota

A study was designed to determine whether the effectiveness of programmed material presented by teaching machine and programmed text varies with Kuder interest pattern. Specifically, 30 introductory psychology students with high mechanical and low literary interest and 30 low mechanical and high literary interest students learned 2 programs. ½ of each group learned the programs via teaching machine while the remaining halves were instructed by programmed text. While no interest patterns or mode of presentation effects were uncovered, a significant Interests × Mode interaction appeared for the longer of the 2 programs. The implications of the results for educational and industrial training programs are discussed.

Studies which have compared programmed texts and teaching machines as modes of presenting programmed material generally find few differences in acquisition and retention differentiating the two approaches (e.g., Eigen, 1962; Goldberg, Dawson, & Barrett, 1964; McKeown, 1965. Also cf. Goldstein, 1964; Goldstein & Gotkin, 1963). However, it is possible that, while the two modes of presentation might be comparable with groups of unselected *S*s, individuals might respond differentially to the two procedures as a function of personality/interest pattern differences. In fact, a related suggestion has been made by Ripple, O'Reilly, Wightman, and Dacey (1965), who feel that personality characteristics of the learner may differentially affect learning under programmed instruction and conventional teaching regimens.

A particularly intriguing hypothesis suggests that *S*s may learn differentially from teaching machines and programmed texts as a function of interest differences. For example, *S*s with high mechanical interest might be expected to learn better from teaching machines than persons with high literary interest, while the latter group might "outlearn" the former on material presented by programmed text. The present study was designed to determine if interest patterns reflecting literary or mechanical preferences are related to the differential effectiveness of text- and machine-presented programmed materials.

METHOD

Subjects

The *S*s were selected from 93 male volunteers drawn from two introductory psychology courses at St. Cloud State College. All of the students were administered the Kuder Preference Record, Form CH. On the basis of these scores, two groups were formed: Group L_{hi}–M_{lo} (above the total sample median on the Literary scale and below the total sample median on the Mechanical scale) and Group M_{hi}–L_{lo} (below the total sample median on the Literary scale and above the total sample median on the Mechanical scale). Those *S*s who did not conform to these combinations of scores were eliminated from the experiment at this point. Also, because the statistical design employed required cell-to-cell proportionality, some additional *S*s were randomly eliminated and the study was run with 30 M_{hi}–L_{lo} and 30 L_{hi}–M_{lo} *S*s, half of each sample being assigned to a teaching machine group and half to a programmed text group. Minnesota Scholastic Aptitude Text scores were available on 51 of the 60 *S*s, permitting a rough evaluation of the intellectual abilities of the four resulting samples. An *F* test run between the means of the four samples proved to be nonsignificant ($F = 2.09$, $df = 3/47$; $M = 34.0$, $SD = 14.1$); it appears likely, therefore, that intellectual differences did not materially affect the results.

[1] The authors are indebted to Vi Kohorst and Kate Tidd for their assistance in the preparation of this paper.

Programs

Two short linear programs were prepared especially for the project.[2] One, arranged in a fading format and consisting of 19 frames, was designed to lead to *memorization* of the following fictitious business address:

Mr. Karl G. Lunden
Chief of Product Development
314 Brewster Building
101 Hollyhock Drive
Mesa, California

The second program was intended to *teach the principle* underlying the digit-letter sequence:

9 D 7 H 5 N 3 T 1

This program consisted of 21 frames with overlapping stimuli. Both programs required constructed responses. The machine presentation was implemented by means of a MTA-100 Scholar teaching machine. The special booklets prepared for the text presentation were arranged in a horizontal format.

Procedure

The 30 *S*s each in the L_{hi}–M_{lo} and the M_{hi}–L_{lo} groups were assigned randomly to either a machine or a text treatment condition. Upon arrival at the experimental room each *S* was exposed to printed instructions regarding the use of the relevant presentation mode. These instructions were comparable except for necessary references to either the machine or the text. While the machine program was administered individually, the booklet was administered to students in groups of about 15. The *S*s were given all the time needed to complete the experimental programs assigned to them. (The average time required for the program was about 15–20 min.) After completing the program *S*s were excused with the explanation that they would be called back at the same time the following day. No description of the content of the second session was offered at this point. The next day *S*s were examined for retention of the programmed material. This involved instructions to reproduce the business address and the digit-letter series learned on the preceding day. The dependent variable used for analysis was the number of individual *characters* correctly recalled. Thus, in the business address there was a total of 88 letters and numbers, and a perfect score necessitated the recall of all characters in the correct order. Correspondingly, a maximum of nine characters could be recalled in the digit-letter sequence.

RESULTS

Business Address Retention

The means of the total number of business address characters recalled after 24 hr. are summarized in Table 1. Differences among the means of the rows and columns were tested for significance at the .05 level by a two-day analysis of variance. Table 2 summarizes this analysis.

TABLE 1

MEANS FOR TOTAL NUMBER OF BUSINESS ADDRESS AND DIGIT-LETTER CHARACTERS RECALLED AFTER 24 HR.

Group	Machine	Text	Total
L_{hi}–M_{lo}[a]			
Address	50.00	62.80	56.40
Digit-letter	7.28	6.67	6.98
M_{hi}–L_{lo}[b]			
Address	66.74	51.07	58.90
Digit-letter	7.87	7.87	7.87
Total			
Address	58.37	56.94	57.66
Digit-letter	7.58	7.27	7.42

Note.—$n = 15$ for each cell.
[a] Above the total sample median on the Literary scale and below the total sample median on the Mechanical scale.
[b] Below the total sample median on the Literary scale and above the total sample median on the Mechanical scale.

The implications of these data are as follows:

1. It is obvious that for the combined samples there is no significant difference between the effectiveness of the programmed text and the teaching machine.

2. When data from the programmed text and the teaching machine are combined there is no significant difference between the performances of the two interest groups.

3. A significant interaction is observed between interest pattern and mode of presentation. By comparison with the programmed texts, the teaching machine appears to be a more appropriate technique for those with high mechanical and low literary interests

TABLE 2

ANALYSIS OF BUSINESS ADDRESS RETENTION

Source	*SS*	*df*	*MS*	*F*
Machine–Text	30.8	1	30.8	<1.00
Group	93.7	1	93.7	<1.00
Interaction	3038.8	1	3038.8	4.02*
Within cells	42294.3	56	755.26	
Total	45457.6	59		

* $p < .05$.

[2] Copies of the programs are available from the senior author upon request.

while the opposite is true for those with low mechanical and high literary interests.

Digit-Letter Retention

The means for the four samples on the digit-letter task are also summarized in Table 1.

The digit-letter retention scores were analyzed in the same manner as the business address data in Table 1. However, as neither the differences among marginal means nor the Row × Column interaction approached significance, the analysis is not described in a separate table. The *F*s for interest pattern, mode of presentation, and interaction were 1.50, .17, and .16, respectively, all *df*'s = 1/56.

DISCUSSION

It is evident that with at least some programs (e.g., the business address program) a significant interaction exists between interest pattern and mode of presentation.

A sophisticated theoretical analysis is not needed to account for the differences observed between the L_{hi}–M_{lo} and the M_{hi}–L_{lo} groups. A student who "likes" things mechanical and "dislikes" printed materials might be expected to respond more favorably to the teaching machine on motivational grounds. Conversely, an *S* who "likes" books but "dislikes" machines might be expected to exhibit an opposite effect relative to the two presentation modes.

The results may have particularly broad implications for future development in the programmed instruction field. They suggest that training programs for groups with high mechanical interests (presumably factory workers, tradesmen, etc.) might best be structured for machine presentation while those designed for persons with stronger literary interests (e.g., students, clerical workers) are probably best presented as programmed texts. Additional research on the relationships between interests and differential effectiveness of various types of programs seems to be very much in order.

As has been noted, no interaction effect was observed with the digit-letter program. This variable influence of interest patterns upon retention may be referable to specific differences in the two programs. First, a fundamental difference in format is to be noted. The business address program employed a fading format in which the general instructional goal is memorization. The digit-letter program, on the other hand, was designed to teach a particular principle from which the construction of a particular series of characters could follow.

Second, the difference in the number of characters in the two programs is to be noted. While a maximum of 88 correct responses were involved in the retention of the business address, the digit-letter series included only 9 characters. Thus, the variability in recall was restricted in the latter program due to a much lower maximum "ceiling" of retention.

In any event it would appear that with some programmed materials arranged in certain formats, retention varies as a function of particular patterns of measured interests. The more specific nature of this interaction awaits further analysis as does the relationship between interest and personality variables, and criteria other than retention.

REFERENCES

EIGEN, L. D. A comparison of 3 modes of presenting a programmed instruction sequence. *Journal of Educational Research,* 1962, **55,** 453–460.

GOLDBERG, M. H., DAWSON, R. I., & BARRETT, R. S. Comparison of programed and conventional instruction methods. *Journal of Applied Psychology,* 1964, **48,** 110–114.

GOLDSTEIN, L. S. Research in programmed instruction: An overview. In G. D. Ofiesh & W. C. Meierhenry (Eds.), *Trends in programed instruction.* Washington: National Education Association, 1964.

GOLDSTEIN, L. S., & GOTKIN, L. G. A review of research: Teaching machines vs. programed textbooks as presentation modes. *Journal of Programed Instruction,* 1963, **1,** 29–37.

MCKEOWN, E. N. A comparison of the teaching of arithmetic in grade four by teaching machine, programmed booklet, and traditional method. *Ontario Journal of Educational Research,* 1965, **7,** 289–295.

RIPPLE, R. E., O'REILLY, R., WIGHTMAN, L., & DACEY, J. Programmed instruction and learner characteristics: Preliminary data. *Psychological Reports,* 1965, **17,** 633–634.

(Received December 26, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 6, 457–461

PERCEIVED DEFICIENCIES IN INDIVIDUAL NEED FULFILLMENT OF CAREER ARMY OFFICERS

PAUL V. JOHNSON [1] AND ROBERT H. MARCRUM [2]

Purdue University

This research studied the extent of individual need fulfillment of career United States Army officers at 3 different levels in the organizational hierarchy. The relative importance which the officers attach to the needs was also investigated. A sample of 504 officers in the ranks of captain through colonel responded to a mailed questionnaire. Conclusions are: (*a*) Perceived deficiencies in need fulfillment tend to increase at successively lower levels in the organizational hierarchy. (*b*) In general, need-fulfillment deficiencies are greatest for the needs in the ego and self-actualization categories. (*c*) The deficiencies which appear to have the most deleterious effect on the accomplishment of Army goals are those associated with the opportunity to exercise authority and the comparative amount of recognition received in command and staff positions.

Like many industrial organizations, the United States Army devotes considerable time and effort to the problem of recruitment and retention of qualified managerial (officer) personnel. Another problem of equal importance, which has received perhaps less attention, is that of obtaining maximum effort and value from those officer personnel who have committed themselves to a military career. Myers (1964) and others provide evidence to suggest that certain support factors such as good working conditions, fringe benefits, and fair practices may reduce dissatisfaction to the extent that individuals will stay with the organization, but that a different group of factors causes managers to be highly motivated toward their work. Myers believes that this latter group, which he calls "motivators," includes the factors of achievement, growth, responsibility, and recognition.

McGregor (1957) interprets this same motivational problem in terms of the hierarchy of needs developed by Maslow (1954). He feels that organizations continue to concern themselves with satisfying the physiological and safety needs of their employees through the provision of better wages, better working conditions, and more fringe benefits even after these two lower level needs have been quite well satisfied. The greatest mistake, McGregor believes, stems from management's failure to recognize that employee dissatisfaction and failure to put forth more than minimum effort derive from the lack of opportunities to satisfy their social, ego, and self-fulfillment needs. The needs contained within these three higher level categories, particularly the ego and self-fulfillment needs, closely parallel Myers' "motivators."

In a series of studies of industrial management personnel, Porter (1961, 1962, 1963a, 1963b, 1964) assessed the extent of perceived deficiencies in need fulfillment as a function of job level and of line versus staff type of job. Using questionnaire items classifiable into a Maslow-type need hierarchy system he found that the largest deficiencies existed in the two highest-order need categories and that deficiencies tended to increase at each successively lower level of the management hierarchy. He also found that line managers perceived greater need fulfillment than staff managers but did not differ in regard to the importance they attached to each type need.

PURPOSE

The purpose of the present study was to examine the extent of individual need fulfillment of regular United States Army officers at three different levels in the organizational hierarchy. Deficiencies in need fulfillment were assessed by determining to what extent an incentive is provided for each need versus the extent to which an incentive should be

[1] Associate professor of economics and chairman, Industrial Relations Committee, Purdue University.

[2] Currently assigned as assistant research officer, Office of Military Psychology and Leadership, United States Military Academy, West Point, New York.

provided for each need. The discrepancy between what is provided versus what should be provided is an indirect measure of the satisfaction or dissatisfaction with each need. An investigation was also made of the relative importance which the officers attach to the needs used in the study.

METHOD

Subjects

The *Ss* were 504 regular Army officers who responded to a mailed questionnaire (87.4% response rate). Officers attending Army service schools at three different levels were selected for the sample. The study was limited to officers who expressed their commitment to at least a 20-yr. Army career and to officers in the three "combat" branches, that is, Armor, Artillery, and Infantry. Officers in these branches experience a similar and reasonably clear-cut career pattern of alternating command and staff assignments up through the organizational hierarchy. Table 1 contains additional information pertaining to the sample.

Questionnaire

Using Maslow's need hierarchy as a guide, the needs shown in Table 2 were selected for investigation.

Interest in the extent of perceived deficiency in need fulfillment prompted the use of two similarly constructed questions for each need. The first question asked essentially, "How much opportunity *do you have* at your current rank to satisfy a particular need?" The second question asked, "How much opportunity *should you have* to satisfy this same need?" The needs were divided into two categories due to the types of responses considered appropriate for each category. The first category consisted of those needs where the opportunity for fulfillment in the Army could be reasonably compared with the opportunity for fulfillment provided by other careers for which the officers might quality. These needs included financial security, association and lasting friendships with high-caliber persons, independent thought and action, prestige, realizing full potential, and a feeling of worthwhile accomplishment. The first question for each of these needs, then, asked the officers to compare need fulfillment opportunities in the Army to other careers with responses on a 5-point Likert scale of less to more. The second question for each need was similarly constructed except for dropping off the lowest response (less) and adding a response to the upper end (much more). This was done in an effort to obtain a more normal distribution of responses to the second question. Since it was unlikely that a respondent would state that he should have less opportunity to satisfy a particular need than he is now being provided, identical scales for both questions would tend to force the answers to the second question into the upper end of the scale. The second category of needs consisted of those which had an internal orientation and did not lend themselves to a comparative approach. These needs included those for authority and for recognition in command and staff positions. Questions concerning these needs were stated in terms of how much recognition *do you* and *should you* receive and how much authority *do you* and *should you* get to exercise, considering the amount of responsibility associated with the job. This dictated the use of a low-high scale with responses to the first question for each need of low, somewhat low, neither especially high nor low, somewhat high, and high. As with the first category of questions, the responses to the second question were altered by dropping off the lowest response (low) and adding a response to the upper end (very high). An open-ended question was used to solicit any comments the *Ss* desired to make regarding their responses to the two questions pertaining to each need. Finally, *Ss* were asked to rank in order of importance to them the specific needs under study. The questionnaire was pretested on 20 locally available officers.

TABLE 1

OFFICERS SAMPLED BY RANK CLASS

Item	Captains	Majors	Lt. Colonels/ Colonels
n	271	164	69
Age			
M	27.98	35.42	44.12
SD	2.10	2.00	2.72
Years service			
M	5.65	13.00	21.67
SD	1.41	1.46	2.18

Note.—$N = 504$.

TABLE 2

SPECIFIC NEEDS BY NEED CATEGORY INCORPORATED INTO QUESTIONNAIRE

Need category	Specific needs investigated: Need for
Safety	Financial security
Social	Association and lasting friendships with high-caliber persons
Ego	
Self-esteem	Authority
	Independent thought and action
Reputation	Prestige (from people outside the military)
	Recognition (credit for accomplishment) in command positions
	Recognition (credit for accomplishment) in staff positions
Self-actualization	Realizing full potential
	Feeling of worthwhile accomplishment

RESULTS

Perceived need-fulfillment deficiencies were measured by computing the difference between responses to the two questions for each need. These difference scores were analyzed and the data are presented in two different but related contexts. In the first place, there is interest in determining the relative amount of dissatisfaction among the various needs studied. Table 3 lists the mean difference scores for each need by rank class. The absolute values of these scores are not as important as the relative values from one need to another. Furthermore, certain trends are apparent when the mean difference scores for the same need are compared across rank classes.

It can be seen from Table 3 that the captains are most dissatisfied with the prestige which is accorded their rank by people outside the military while the majors and colonels are most dissatisfied with the amount of authority they are permitted to exercise in command positions. It is interesting to note that the captains have three scores which are higher than the maximum score for the majors. On the other hand, the highest score in the group is that of the colonels' score on authority. With six of the nine needs, the captains' score is higher than the majors' score and the majors' score is higher than the colonels' score. This suggests, of course, that need-fulfillment opportunities are better at successively higher levels in the organizational hierarchy. In general, the greatest deficiencies exist in the ego and self-actualization categories.

TABLE 3

MEAN DIFFERENCE SCORES BY RANK CLASS AS A MEASURE OF THE EXTENT OF PERCEIVED DEFICIENCIES IN NEED FULFILLMENT

Need for	*M* difference scores		
	Captains	Majors	Colonels
Financial security	1.3727	1.3780	1.3768
Association and lasting friendships with high-caliber persons	.7085	.5244	.3043
Authority	1.8044	1.6280	1.9420
Independent thought and action	1.8155	1.5061	1.3188
Prestige	1.8893	1.3561	1.3188
Recognition (command positions)	1.4502	1.3598	1.2029
Recognition (staff positions)	.8930	.6463	.7391
Realize full potential	1.6125	1.5000	1.1884
Feeling of worthwhile accomplishment	1.0959	.8049	.6667

Though the trends in mean difference scores among the rank classes have been noted, it was considered desirable to determine whether there were statistically significant differences in the distribution of difference scores among rank classes for each need. Rather than using the mean scores and

TABLE 4

SIGNIFICANT DIFFERENCES AMONG RANK CLASSES IN THE EXTENT OF PERCEIVED DEFICIENCIES IN NEED FULFILLMENT

Need	Kruskal-Wallis *H* values			
	Overall	Capt-Maj	Maj-Col	Capt-Col
Financial security	*ns*	—	—	—
Association and lasting friendships with high-caliber persons	9.5835*	4.2970*	*ns*	8.0991*
Authority	*ns*	—	—	—
Independent thought and action	9.3400*	5.1252*	*ns*	6.8548*
Prestige	20.8152*	18.0205*	*ns*	7.7898*
Recognition (command positions)	*ns*	—	—	—
Recognition (staff positions)	*ns*	—	—	—
Realize full potential	*ns*	—	—	—
Feeling of worthwhile accomplishment	7.0665*	3.6909*	*ns*	5.4324*

* $p < .05$.

some parametric statistic wherein the ability to meet the assumptions inherent in such a statistic would be questionable, the Kruskal-Wallis *H* test was used. Hays (1963) believes that this is a good rank-order method of determining whether a group of independent samples comes from the same or different populations and states that it compares very well with the *F* test from the analysis of variance. Hays (1963) points out that once the *H* has been computed, a correction formula should be applied in any case where the number of *Ss* is small or where there is a large number of tied observations. Since there was a substantial number of tied observations in the distributions of difference scores, this correction is applicable here. The correction serves to increase the size of the *H*, so a correction was applied when the matter of significance was in question.

Table 4 shows the results of these computations. An overall *H* was computed for each need and, where the *H* was significant, three additional *H* tests were computed to determine the difference between any two rank classes. It can be seen that, for the four needs where the overall *H* tests are significant, in each case there are significant differences between the captains and majors and the captains and colonels but not between the majors and colonels.

The results of the investigation of the relative importance which *Ss* attach to the needs used in the study are contained in the following hierarchy of needs going from most important to least important:

1. Feeling of worthwhile accomplishment.
2. Opportunity to realize full potential.
3. Opportunity for independent thought and action.
4. Financial security.
5. Opportunity to associate with high-caliber persons.
6. Recognition.
7. Opportunity to exercise authority.
8. Prestige.

No differentiation was made between recognition in command and staff positions. Based on an examination of the frequency distribution of responses for each need, the median response was selected as the most meaningful index of central tendency for the purpose of determining the rank order. For simplicity, only the rank order based on the entire sample is presented. When the rank classes were taken separately, there were slight and generally insignificant differences in the rank order for majors and colonels. The greatest consensus among all *Ss* existed for the two most important and two least important needs.

DISCUSSION

The Army, like the industrial organizations studied by Porter, has apparently created an environment in which the opportunities for higher level need fulfillment are greater at successively higher levels in the organizational hierarchy. Such an environment is conducive to continued growth and development of managerial (officer) personnel, and it would appear that guidance of junior officers should stress the ever-increasing opportunity for higher level need satisfaction as they progress through their careers. On the other hand, need-fulfillment opportunities at the lower managerial levels must be sufficient to prevent widespread dissatisfaction at these levels.

Some interesting implications emerge when the obtained need-fulfillment deficiencies are considered in conjunction with the relative importance which *Ss* attached to the needs. Based on Maslow's hierarchy it would be predicted that needs from the ego and self-actualization categories would be the most important and perceived as the least fulfilled. This prediction generally holds up for the three needs perceived as most important but is inconsistent with the results pertaining to the needs for prestige and for the opportunity to exercise authority. Not only are these needs rated in the last and next to last position by all ranks, but also the officers perceive relatively large need-fulfillment deficiencies for these needs. This raises the question of whether these needs are truly unimportant to the officers studied, relative to the other needs, or whether the extent of perceived need-fulfillment deficiencies is sufficiently large to have caused the officers to downgrade the importance of these needs because they see no hope of satisfying them.

Regardless of the reason for the need for

authority receiving such a low ranking, the findings relative to this need appear to present the Army with a problem of considerable magnitude. The Army, of necessity, must place great importance on the development of commanders. A Department of the Army (1964) circular states that "command and supervisory positions enjoy the highest prestige of all positions in the Army, and therefore are the most sought after duty assignments [p. 8]." The low ranking and the dissatisfaction with the opportunity to exercise authority do not bear out this statement. The majors and colonels expressed their greatest dissatisfaction with this need. Not only is it a reversal of trend for the colonels to express more dissatisfaction with this need than the two lower ranks, but their score is the highest of all dissatisfaction scores in the study. The questions on authority elicited a number of open-ended responses which may help to explain these results. The following quote from one lieutenant colonel expresses the general tenor of these responses:

> One factor, more than any other, promises to affect adversely motivation. That is, the trend toward highly centralized control and management of all military endeavor. This trend not only will systematically reduce the aggressive initiative of those already in the Army but also will reduce the interest of younger men of high caliber who might otherwise have been attracted to the challenges of military life. At its worst, the trend could produce a mediocre Army of conformists, input data statisticians, and escapists looking for immunity from the responsibilities of decision.

Though the Army relies heavily on the proper exercise of authority, it would appear that the conditions under which such authority is being exercised are less than favorable.

A related problem stems from the findings relative to the need for recognition or credit for accomplishment in command and staff positions. The data show considerably more dissatisfaction with the amount of recognition received in command positions than in staff positions which suggests that the amount of recognition being given is not commensurate with the responsibility associated with each type position. This state of affairs can only serve to detract from the Army's emphasis on the attractiveness of command positions. Considering these results in conjunction with the results concerning the need to exercise authority, it would not be unlikely that officers would lean toward staff positions where the risk of failure is much less than that for command positions.

The respondents' reactions to the questions related to the feeling of worthwhile accomplishment deserve some comment. The officers approached unanimity in the selection of this need as the most important, and the perceived deficiency in fulfillment of this need was not as great as might have been predicted. The desire to explain their reasons for selecting this need as most important prompted more open-ended responses than any other item. In general, the officers consider that the Army has a mission of great significance, and the opportunity to serve one's country in such a capacity is the most important reward of an Army career. They also suggest that the magnitude of this reward is sufficient to overcome some of the disadvantages inherent in a service career.

REFERENCES

DEPARTMENT OF THE ARMY. Career planning for Army officers. Pamphlet No. 600-3, June 1964.

HAYS, W. L. *Statistics for psychologists*. New York: Holt, Rinehart & Winston, 1963.

MASLOW, A. H. *Motivation and personality*. New York: Harper, 1954.

MCGREGOR, D. M. The human side of enterprise. *Management Review*, 1957, **46**, 22–29, 88–92.

MYERS, M. S. Who are your motivated workers? *Harvard Business Review*, 1964, **42**, 73–88.

PORTER, L. W. A study of perceived need satisfactions in bottom and middle management jobs. *Journal of Applied Psychology*, 1961, **45**, 1–10.

PORTER, L. W. Job attitudes in management: I. Perceived deficiencies in need fulfillment as a function of job level. *Journal of Applied Psychology*, 1962, **46**, 375–384.

PORTER, L. W. Job attitudes in management: II. Perceived importance of needs as a function of job level. *Journal of Applied Psychology*, 1963, **47**, 141–148. (a)

PORTER, L. W. Job attitudes in management: III. Perceived deficiencies in need fulfillment as a function of line versus staff type of job. *Journal of Applied Psychology*, 1963, **47**, 267, 275. (b)

PORTER, L. W. Job attitudes in management: IV. Perceptions of the importance of certain personality traits as a function of job level. *Journal of Applied Psychology*, 1964, **48**, 31–36.

(Received December 26, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 6, 462–468

A CORRELATIONAL-CAUSAL ANALYSIS OF THE RELATIONSHIP BETWEEN EXPECTANCY ATTITUDES AND JOB PERFORMANCE

EDWARD E. LAWLER, III [1]

Department of Administrative Sciences, Yale University

The causal basis for the relationship between expectancy attitudes and job performance was tested in this study by the use of 2 correlational approaches to determining causality. Data were gathered from 55 managers on the probability that certain rewards could be obtained as a result of their job performance. In addition, self, superior, and peer ratings of job performance were obtained for each manager. Similar attitude and performance data were collected at 2 points in time separated by about 1 yr. A cross-lagged correlational analysis of the data generally supported the view that the expectancy attitudes could be best thought of as causing the job performance. A dynamic correlational analysis of the data failed to show strong positive results. The implications of these findings for the time lags that are involved in expectancy attitudes causing job performance were considered.

A number of recent studies have demonstrated that a significant relationship exists between employees' beliefs about the probability that rewards depend upon their job behavior and their job performance. Lawler and Porter (1967a), for example, have found that the more managers believe rewards like pay, promotion, and respect from others stem from good performance the more likely they are to be rated as good performers by their peers, superiors, and selves. Generally, similar findings have been reported in four other studies of expectancy, reward-probability, or path-goal attitudes as this kind of belief has been referred to by different researchers (Galbraith & Cummings, 1967; Georgopoulos, Mahoney, & Jones, 1957; Porter & Lawler, 1968; Spitzer, 1964). These studies have also tended to show that when expectancy attitudes are weighted by the value or importance of the rewards considered, the correlations between the attitudes and job performance are increased. Lawler and Porter (1967a), for example, report a correlation of .25 between unweighted expectancy-attitude items and a measure of performance, but a correlation of .38 between performance and the same expectancy-attitude items when they had been weighted by the value of the rewards to the individual managers.

The finding that expectancy attitudes are significantly related to performance is, of course, predictable from the kind of expectancy theories of motivation that have recently been stated by Vroom (1964), Atkinson (1964), and others. Earlier statements of expectancy theory by Peak (1955) and Rotter (1955) also are congruent with these findings. All these statements of expectancy theory predict that a positive relationship between expectancy attitudes and performance should exist because expectancy attitudes are reflections of the individual's motivation to perform. Thus, the expectancy attitudes are essentially seen to be causing or leading to the performance. Hence, the relationship is expected not because the performance causes the expectancy attitudes or because a third variable causes both but because the expectancy attitudes cause the performance.

All of the previous studies of the relationship between expectancy attitudes and performance have used a static correlational approach and thus it is impossible to determine the causal basis for the relationships found. It typically has been assumed by the researchers that the relationship found came about because the attitudes caused the performance but this assumption has not been tested. As Porter and Lawler (1968) have pointed out, an explanation to the effect that the attitudes caused the performance allows for the most parsimonious explanation of all

[1] The work on this project was done while the author was a Ford Foundation Faculty Research Fellow.

the data collected and it fits with motivation theory. However, before it is safe to assume that the causal direction runs from the attitudes to the performance, direct tests of the causal direction are needed. The present study represents an attempt to fill this gap in the literature by specifically determining the causal basis for the commonly found relationship between expectancy attitudes and job performance.

The usual way of determining whether two variables are causally related is, of course, to do an experiment in which one variable can be manipulated in order to observe its effect on another. There are, however, certain research problems that do not lend themselves to the experimental model regardless of whether a field or laboratory experiment is done. Experimental laboratory studies often do not seem to offer an adequate enough simulation of the real world for results to be used with confidence when more applied problems are being dealt with, while field experiments present a number of practical and ethical problems. These limitations of the experimental approach are certainly present when consideration is given to testing for the causal relationship between expectancy attitudes and job performance. There are, however, two alternatives to the traditional experimental approach that would seem to offer ways of looking at the causal basis for the relationship considered in this paper. Both of these approaches can be classified as quasi-experimental or correlational-causal in nature since they allow causal inferences to be made with some degree of confidence without the usual experimenter (E) manipulation of variables or E control over the situation being present.

The first of these approaches is the cross-lagged panel correlational method. This approach was initially discussed by Simon (1954) and has been more recently considered by Blalock (1962), Pelz and Andrews (1964), and Campbell and Stanley (1963). It involves gathering similar data on the two variables that are to be tested for causality at two different points in time. In the case of the present problem, this would mean gathering identical expectancy-attitude data and identical job-performance data at two separate points in time. Once these data have been collected, the kind of cross-lagged correlational analysis shown in Figure 1 can be set up to analyze the relationship between expectancy attitudes and job performance.

This analysis yields six correlation coefficients, two of which (3 and 4) tell something about the stability of the measures and are not directly involved in tests for causality, while the others can give some indication of the causal direction of the relationship found between the attitude and performance measures. If the remaining correlations shown in Figure 1 stand in the following relationship to each other: $5 > (1 = 2) > 6$, then there is good reason to believe that the expectancy attitudes caused the performance. However, if the correlation coefficients stand in this relationship: $6 > (1 = 2) > 5$, then there is reason to believe that the performance caused the expectancy attitudes. The one weakness in this approach is that it cannot rule out the possibility that a third variable causes the two other variables to covary.

The logic underlying this kind of analysis rests upon the time lag that typically exists when one variable causes another. The argument is that if A causes B then the present state of A should be more strongly related to B's future state than to B's past or present state. Thus, where $A \rightarrow B$, then r_{AB} where B is measured after A should be greater than r_{AB} where B is measured either before or at the same time as A. Thus, by comparing the relative sizes of r_{AB_1}, r_{AB_2}, and r_{AB_3} where B is measured before, after, and at the same time as A, it is possible to determine whether the hypothesis $A \rightarrow B$ or $B \rightarrow A$ is more tenable. Translating this to the present problem it means that, if the expectancy attitudes

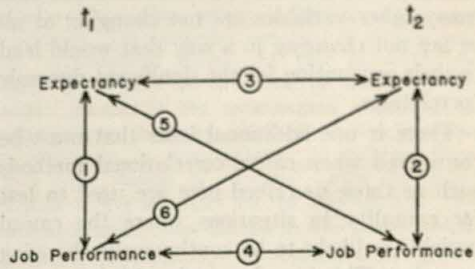


FIG. 1. Cross-lagged correlational model.

cause the performance, the strongest relationship should be between the attitude measure at Time 1 (t_1) and the performance measure at Time 2 (t_2).

Vroom (1966) has recently suggested a second method that can be used in order to determine the causal basis for the relationship between two variables. This method, like the cross-lagged approach, requires identical measurements at two points in time. It has the advantage of being strong where the cross-lagged approach is weak, that is, it can help rule out the possibility that a third variable is responsible for the relationship found. However, unlike the cross-lagged approach it cannot tell whether $A \rightarrow B$ or $B \rightarrow A$. All it can tell is something about whether a third variable has caused both. Thus these two approaches are complementary in nature and should be used together.

Dynamic correlations are computed by correlating the difference between A at t_1 and t_2 with the difference between B at t_1 and t_2. Thus, in effect, changes in one variable are being correlated with changes in another variable. To produce a spurious dynamic correlation between two different scores ΔA and ΔB, a third variable must change in different amounts or directions in the members of the sample, and these changes must be highly correlated with both ΔA and ΔB. Obviously it is extremely unlikely that any variable will meet these conditions and, because of this, most potential third causal variables can be ruled out. Thus when a significant dynamic correlation is found between two variables, one can put more faith in the fact that one caused the other than one can when a significant static correlation appears between two variables. This is particularly true in situations where the researcher can determine that many other variables are not changing at all or are not changing in a way that would lead to their accounting for the significant dynamic correlations.

There is one additional issue that must be considered when causal-correlational methods such as those described here are used to test for causality in situations where the causal variable is likely to be continuously changing over time. This issue has to do with how much time actually elapses between changes in one variable and changes in the other. This issue also serves to highlight a difference between the two methods. Dynamic correlations would seem to be most likely to show positive results when the change in one variable leads almost immediately to a change in the other variable, that is, where the two variables can, for practical purposes, be thought of as moving almost together, although causality still exists. Since this method compares changes in A with changes in B between two points in time, the changes in A must be quickly reflected in B if the changes in the two from t_1 to t_2 are going to be highly correlated. If a long time lag exists between changes in A and their effects on B then the changes in A between t_1 and t_2 will not be reflected in the changes in B between t_1 and t_2 regardless of how far t_1 and t_2 are separated. They would be more likely to appear in the difference between B at t_2 and t_3.

The cross-lagged approach would seem to be most likely to show positive results when there is some delay between changes in the first variable and changes in the second variable, and when the separation between data-collection points approximates the delay between changes in one and changes in another. If the impact of the change is very rapid, then the cross-lagged approach will not show positive results since it depends on Variable A measured at t_1 influencing Variable B measured at t_2 more strongly than it influences Variable B measured at t_1. If the time lag for the effect of A on B is very short the effect probably will already have taken place at t_1. And, thus, A will be as highly correlated with B measured at t_1 as with B measured at t_2. Considering the specific problem in the present study, it would seem that there may be considerable delay between changes in expectancy attitudes and changes in performance. Although an individual might start working hard immediately after an attitude change it undoubtedly would take a period of time for this actually to show up in the relatively crude performance measures that are presently available. On the other hand, if t_1 and t_2 are not separated by a sufficient period for A at t_1 to affect B at t_2 the cross-lagged analysis will not show positive results. In order to allow the cross-lagged

approach sufficient time to show positive results, it was decided to separate the two data-collection sessions by 1 yr.

METHOD

Sample

Data were collected from 55 managers in social service agencies. Included were the directors and all the managers of a YMCA, a visiting nurse service, a Girl Scout council, and a city department of health. Of the 55 managers studied, 41 could be considered first-level supervisors while the remaining managers were at higher levels of supervision. A comparison between the demographic characteristics of the sample and the characteristics of other samples of middle- and lower-level managers (e.g., Porter, 1961, 1962) revealed good comparability with one exception. Thirty-one of the managers in the present sample were females while none of the previous samples contained any female managers. Other studies (Lawler, 1967; Lawler & Porter, 1967a, 1967b) have included data gathered from part of this sample.

Attitude Measures

The attitude data were gathered by means of a short questionnaire. One part of the questionnaire was designed to measure the managers' reward-expectancy attitudes. It listed six rewards each of which was assumed to be related to the satisfaction of one or more needs: pay, promotion, prestige, security, autonomy, and opportunities to use skills and abilities. For each of these six rewards the managers were asked to indicate on a 7-point scale how helpful six factors were for obtaining the rewards in their organizations (see Lawler & Porter, 1967a, 1967b, for a description of this questionnaire). Two of the factors that were listed were particularly concerned with the job performance of the managers (working hard and quality of job performance). Previous studies (Lawler & Porter, 1967a) have shown that these two items tend to correlate highly for managers when the managers are asked to indicate how helpful the items are in obtaining important rewards.

A second part of the questionnaire contained a section that asked each manager to indicate how important the six reward items were to him. For these ratings a 7-point scale (1 unimportant to 7 very important) was used. In order to develop expectancy-attitude measures for each manager, the data from the importance measures were combined with the data concerned with the likelihood that good performance and working hard would lead to the rewards. As suggested by the expectancy theories of motivation, this was done by multiplying the importance of the reward times the degree to which it was seen to be obtainable from both good performance and working hard. The results of these two multiplications were then summed for each of the six rewards so that for each manager a composite expectancy-attitude measure was developed of the degree to which he felt receiving each of the six rewards was based upon his job performance.

Job-Performance Measures

Each manager was ranked relative to other managers holding similar jobs by his superior and by his peers. Managers were ranked on four factors, one of which, the quality of his job performance, is relevant for the present study. The rankings were converted to standard scores with a mean of 50 and a standard deviation of 10. Where a manager was ranked by more than one peer, an average peers' ranking was computed. In addition to being ranked by his superior and peers, each manager was asked to rate himself on a number of factors, one of which was the quality of his job performance. These ratings were made on a 7-point scale ranging from 1 (low in comparison to similar others) to 7 (high in comparison to similar others).

Analysis of the rankings showed fairly good agreement between the superior and peer rankings (e.g., $r = .56$ between peer and superior rankings on job performance at t_1 and $r = .54$ for t_2). It was encouraging to note that superiors' and peers' rankings correlated most highly when they were ranking the managers on identical factors. The correlations between superior and peer rankings on the same factor also generally tended to be higher than the relationships among the superiors' and peers' rankings of the same people on different factors. Thus, the superiors' and the peers' rankings substantially met the requirements of convergent and discriminant validity (Campbell & Fiske, 1959). The situation with respect to the self ratings was quite different, however. The self ratings of job performance showed only a low relationship to the peers' rankings ($r = .26$ at t_1) and to the superiors' rankings ($r = .09$ at t_1). Thus, the self ratings did not show convergent and discriminant validity.

Procedure

Two meetings, 1 yr. apart, were held with all the participants in the study. At both meetings they were asked to fill out similar attitude questionnaires and ranking forms. It was carefully explained to all participants that they were participating in a university-sponsored research project and that all the data obtained from them would be held in the strictest confidence. The researcher remained in contact with the organizations during the year in which the study took place in order to be aware of any major changes in the organizations. None of the organizations, however, changed in size, purpose, structure, or in administrative procedures. There were some changes in personnel during the period. A total of five managers left the organizations. This, of course, meant some changes in who ranked certain people when the t_2 data were collected. Some managers were ranked by different superiors and peers at t_1 and at t_2.

RESULTS

In presenting the data collected, first the results concerned with the superior rankings will be considered, next those concerned with the peer rankings, and finally those concerned with the self ratings. Figure 2 therefore presents the data for the cross-lagged correlational analysis for the relationships among the expectancy-attitude measures and the superiors' rankings of job performance. Because expectancy theory calls for a linear combination of each manager's six reward-expectancy measures, in order to predict his job performance, multiple correlation coefficients were used for Relationships 1, 2, 5, and 6. These four relationships represent all the instances where attitude measures were correlated with performance measures. Thus, all tests of causality were made by comparing the relative sizes of multiple correlation coefficients. The correlation between the two performance measures is a Pearson product-moment while the correlation between the expectancy attitude measures is the average Pearson r for the six rewards considered.

The correlations presented in Figure 2 between the attitudes and the superiors' performance measures bear the kind of relationship to each other that would be expected if the attitudes caused the performance. The lowest correlation is between the performance at t_1 and the attitudes at t_2, while the highest correlation is between the attitudes at t_1 and the performance at t_2. In addition the correlations between the attitudes and performance data when the coefficients are computed from data collected at the same point in time fall between these two correlations. In order to compute the dynamic correlations between the changes in the attitude variables and the changes in the performance measure, a multiple correlation coefficient was used since changes in six attitude variables were being related to the changes in the performance measure. For the superiors' rankings, this coefficient did show positive Beta weights for the predictors as would be expected if increases in the attitude measure were associated with increases in the performance measure but the coefficient did not reach statistical significance ($R = .32$).

Figure 3 presents the cross-lagged correlational data for the analysis of the peers' rankings of job performance. These data do not provide a close fit for the model that indicates the attitudes caused the performance. The relationship of Correlation 5 to Correlation 6 suggests that the attitudes caused performance, but Correlations 1 and 2 do not fall in between these two as they would be expected to if this analysis were to provide strong support for the view that the attitudes caused the performance. The multiple correlation computed between the changes in the attitude measures and the changes in the performance measures was not significant although the Beta weights were positive ($R = .41$).

Figure 4 presents the cross-lagged correlational data for the analysis of the self ratings of job performance. As was true for the peer ratings, this analysis is suggestive of the fact that the attitudes caused the performance but the fact that Correlations 1 and 2 do not fall between 5 and 6 prevents the data from perfectly fitting the model of the attitudes causing the performance. Nevertheless, the fact that the attitudes were better predictors of the performance in the future than the performance was of the attitudes in the future is

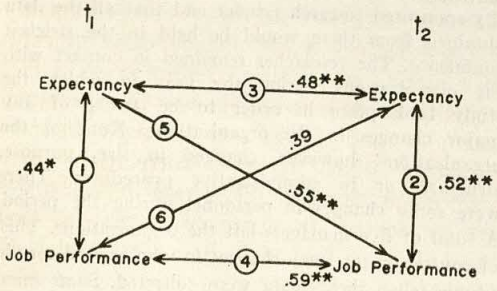


FIG. 2. Cross-lagged correlational analysis for measures of expectancy attitudes and superiors' rankings of job performance (* $p < .05$; ** $p < .01$).

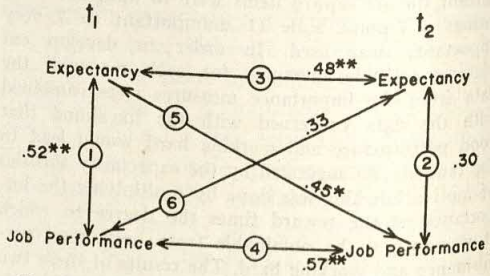


FIG. 3. Cross-lagged correlational analysis for measures of expectancy attitudes and peers' rankings of job performance (* $p < .05$; ** $p < .01$).

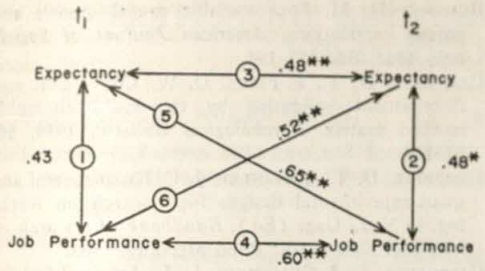


FIG. 4. Cross-lagged correlational analysis for measures of expectancy attitudes and salf-ratings of job performance (* $p < .05$; ** $p < .01$).

suggestive of the fact that the performance caused the attitudes. The dynamic correlation between the changes in the expectancy attitudes and the changes in the performance measure was significant ($R = .54$, $p < .01$), and as was expected the Beta weights were positive.

DISCUSSION

The results of the study are generally supportive of the view that expectancy attitudes can best be thought of as causing performance. The cross-lagged evidence would seem to rule out the possibility that the frequently found relationship between expectancy attitudes and performance comes about because the performance causes the expectancy attitudes. The fact that in all three cross-lagged analyses the attitudes at t_1 predicted the performance at t_2 better than the performance at t_1 predicted the attitudes at t_2 would seem to provide good support for this point. Also supportive of this view is the fact that the attitude to performance correlations were generally higher for the predictions of the attitudes into the future than they were for the static correlations between the attitudes and the performance at the same point in time. These latter results also suggest that one way previous studies might have been able to obtain a stronger expectancy attitude to performance relationship would have been to delay the collection of the performance data until some time after the attitude data had been collected.

The data from the dynamic correlations is only suggestive of the fact that a third variable is not responsible for the relationship between the attitudes and the performance. Only one of the three dynamic correlations was significant, yet all three generally took the form that would be expected if changes in the attitudes led to changes in the performance without a third variable being involved. The data from the cross-lagged analysis suggest that a third variable does not simultaneously cause both the attitudes and the performance, but it does not rule out the possibility that a third variable first causes the attitudes and then the performance, or that the attitudes cause a third variable which in turn causes the performance. Overall, therefore, the data would seem clearly to rule out the possibility that the positive relationship found between performance and expectancy attitudes comes about because the performance causes the attitudes and is suggestive of the fact that it comes about because the attitudes caused the performance.

Part of the reason that the dynamic correlations were not higher may well be found in the time relationships that may occur when expectancy attitudes cause performance. It is reasonable to expect that there might be a fairly long time lag between changes in expectancy attitudes and changes in performance measures. This would seem to be particularly true where superior and peer rankings of performance are used. Even where attitude changes almost directly lead to performance changes, it might be some time before the performance changes could actually be expected to show up in the rankings of the performance. This would be expected because of the time needed to convince the raters that a real change has actually taken place and because of the time needed for the raters actually to become aware of the changes Some support for this thinking is provided by the fact that the dynamic correlation was significant for the self ratings but not for the peer and superior rankings, since presumably the self ratings would change more quickly. If a significant amount of time does in fact actually elapse between changes in expectancy attitudes and consequent changes in performance, then dynamic correlations may never show strong results where these changes are considered. In summary, it may be that the failure to find high dynamic correlations is not so much due to the fact that a third causal variable is involved, but the fact that there is a significant time lag built into the causal relationship.

It is impossible at this time to say exactly what the time sequence actually should be between the two measurements of the attitudes and performance variables if the causal relationship is to be tested most effectively. The 1-yr. separation chosen in the present study was an arbitrary time, hence future research should consider collecting data with differing intervals between the two collection points. A logical next step would appear to be a study where some data are collected with a 6-mo. separation, some are collected with a 1-yr. separation, and some are collected with a 2-yr. separation. By comparing the results obtained for the varying periods of separation it should be possible better to understand the time sequence that is involved in any causal relationship between expectancy attitudes and performance and to test for the causal relationship better. The correlations should be highest at the point in time which most closely approximates the actual time it takes for A to affect B.

Overall, the results obtained from the two correlational-causal approaches that were used are sufficiently interesting to suggest that this approach can be useful in the study of many other topics where causal priorities are in dispute. For example, the relationship between satisfaction and performance could be studied with this approach as could the relationship between participant leadership styles and employee involvement. This, of course, does not mean that this type of analysis can ever replace traditional experiments as the best way to establish causalities since experiments do allow causal questions to be answered with a higher degree of confidence. However, the causal-correlational approaches do have some advantages that argue they deserve more widespread use than they have received. They have some of the advantages of the field experiment without many of the disadvantages. That is, data from which causal inferences can be made are obtained for real situations, as in the field experiment, but many of the ethical, cost, and feasibility problems that go with field experiments are not likely to occur.

REFERENCES

ATKINSON, J. W. *An introduction to motivation.* Princeton: Van Nostrand, 1964.

BLALOCK, H. M. Four variable causal models and partial correlations, *American Journal of Sociology,* 1962, **68**, 182–194.

CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin,* 1959, **56**, 81–105.

CAMPBELL, D. T., & STANLEY, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching.* New York: Rand McNally, 1963.

GALBRAITH, J., & CUMMINGS, L. L. An empirical investigation of the motivational determinants of task performance: Interactive effects between instrumentality-valence and motivation-ability. *Organizational Behavior and Human Performance,* 1967, **2**, 237–257.

GEORGOPOULOS, B. S., MAHONEY, G. M., & JONES, N. W. A path-goal approach to productivity. *Journal of Applied Psychology,* 1957, **41**, 345–353.

LAWLER, E. E., III. The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology,* 1967, **51**, 369–381.

LAWLER, E. E., III, & PORTER, L. W. Antecedent attitudes of effective managerial performance. *Organizational Behavior and Human Performance,* 1967, **2**, 122–142. (a)

LAWLER, E. E., III, & PORTER, L. W. The effect of performance on job satisfaction. *Industrial Relations,* 1967, **7**, 20–28. (b)

PEAK, H. Attitude and motivation. *Nebraska Symposium on Motivation,* 1955. **3**, 149–188.

PELZ, D. C., & ANDREWS, F. M. Detecting causal priorities in panel study data. *American Sociological Review,* 1964, **29**, 836–848.

PORTER, L. W. A study of perceived need satisfactions in bottom and middle management jobs. *Journal of Applied Psychology,* 1961, **45**, 1–10.

PORTER, L. W. Job attitudes in management: I. Perceived deficiencies in need fulfillment as a function of job level. *Journal of Applied Psychology,* 1962, **46**, 375–384.

PORTER, L. W., & LAWLER, E. E., III. *Managerial attitudes and performance.* Homewood, Ill.: Irwin-Dorsey, 1968.

ROTTER, J. B. The role of the psychological situation in determining the direction of human behavior. *Nebraska Symposium on Motivation,* 1955, **3**, 245–268.

SIMON, H. A. Spurious correlation: A causal interpretation. *Journal of the American Statistical Association,* 1954, **49**, 467–479.

SPITZER, M. E. *Goal-attainment, job satisfaction and behavior.* (Doctoral dissertation, New York University) Ann Arbor, Mich.: University Microfilms, 1964. No. 64-10048.

VROOM, V. H. *Work and motivation.* New York: Wiley, 1964.

VROOM, V. H. A comparison of static and dynamic correlation methods in the study of organizations. *Organizational Behavior and Human Performance,* 1966, **1**, 55–70.

(Received December 26, 1967)

Journal of Applied Psychology
1968, Vol. 52, No. 6, 469–473

RELATIONSHIP BETWEEN JOB SATISFACTION AND JOB INVOLVEMENT [1]

PETER WEISSENBERG AND LEOPOLD W. GRUENFELD

Department of Business Enterprise and Accounting, State University of New York at Binghamton

New York State School of Industrial and Labor Relations, Cornell University

This study investigated the relationship between motivator and hygiene satisfaction variables to job involvement. *Ss* were 96 civil service supervisors who completed the Wernimont job-satisfaction scale and a job-involvement measure developed by Lodahl and Kejner. The results of the study were that motivator, but not hygiene, satisfaction variables correlated with job involvement. In addition, total motivator satisfaction scores accounted for considerably more variance in overall job satisfaction than did hygiene variables.

Although job involvement has been related to turnover (Bass, 1965), the determinants of job involvement have not been investigated. It is hypothesized that job satisfaction is one such determinant. The emphasis placed on job-satisfaction variables has resulted in a relative neglect of the job-involvement variable. This is perhaps due to a lack of conceptual differentiation between job satisfaction and job involvement, and to an apparent failure to realize that it is possible for some persons to be highly satisfied, but not involved, and for others to be highly involved, but not satisfied. Moreover, some sources of job satisfaction are probably more likely to be related to job involvement than are others. It was the purpose of this study to investigate the differential relationships between various sources of job satisfaction and job involvement.

Job Involvement

The definition of job involvement is complex. Lodahl and Kejner (1965) consider it to be indicative of an individual's work commitment. Bass (1965) views it as representative of the employee's ego involvement in his job and thus relates it to performance. Bass (1965) finds that the following conditions lead to the strengthening of job involvement: opportunity to make more of the job decisions, the feeling that one is making an important contribution to company success, recognition, achievement, self-determination, and freedom to set one's own work pace. March and Simon (1958) suggest that job involvement may also be related to the decision to participate and the decision to produce. Among other things, both of these decisions are affected by the individual's perception of a variety of alternatives. The person who is more job involved may perceive fewer available alternatives; he may, therefore, be inclined to participate more actively in the required activities of the organization. Katz and Kahn (1966) state: First, that job involvement is a necessary condition if the individual is to accept fully the organizational demands placed upon him by his membership in an organization; second, that the degree of job involvement is related to level of aspiration and to the degree of internalization of organizational goals; and third, that job involvement is a moderator variable in the relationship between satisfaction and performance, for only among involved employees does a positive relationship between satisfaction and performance become evident. Thus job involvement can be considered as an important measure of organizational effectiveness that may be, at least in part, influenced by job satisfaction.

Job Satisfaction and the Two-Factor Theory

Herzberg, Mausner, and Snyderman (1959) formulated the "two-factor theory of job satisfaction." Their research concluded that there is one set of conditions the presence of which induces increased job satisfaction,

[1] This research was supported in part by a United States Public Health Service Fellowship No. 1-F1-MH-33, 376-01 (BEH-A) from the National Institute of Mental Health awarded to the senior author.

but the absence of which does not induce job dissatisfaction; and that there is another set of conditions, the absence of which leads to job dissatisfaction but the presence of which does not lead to increased job satisfaction. They termed the first set of conditions "motivators" since satisfaction with these conditions was "conceptualized as actuating *approach* rather than *avoidance* behavior [Herzberg et al., 1959, p. 114, italics added]." The second set were termed "hygienes" since these conditions appeared to contribute only to the prevention of an "unhealthy psychological work environment [Herzberg et al., 1959, p. 113]." The motivators are: recognition, achievement, advancement, responsibility, and the work itself. These variables have also been termed intrinsic or work-content variables, because they are all presumably derived from performance of the job itself. The hygiene conditions are: interpersonal relations with peers and with superiors, company policy and administration, superior's technical competence, working conditions, and job security. Hygiene conditions have also been referred to as extrinsic or work-context conditions because they are all derived from the environment surrounding the job. Salary, depending upon whether or not it is contingent upon performance, may be either a motivator or a hygiene factor.

Hypothesis

In this study job involvement was considered to be a quasi-indicator of motivation. Therefore, it appeared reasonable to test the Herzberg et al. (1959) generalization using job involvement as the dependent variable. The specific hypothesis which was formulated for this research is as follows: Satisfaction with motivator sources will be related to increased job involvement, but satisfaction with hygiene sources will be unrelated to job involvement.

METHOD

Organization and Subjects

The *Ss* were supervisors of a state (civil service) department normally employing about 5,000 persons. They ranged in rank from first-line supervisors (G 17) to the highest nonappointive position in the department, the executive director (G 38).

One hundred seven persons initially took part in the study. However, all persons did not complete every one of the instruments, and some, although completing each instrument, left some items blank. Female employees were removed from the sample. The final sample consisted of 96 male supervisors who were between 40 and 62 yr. of age. Their salaries ranged from $7,700 to $17,300. They supervised from 0 to 18 persons (some were supervisors in name only). The *Ss* had a range of 8–35 yr. of service. The educational level of the supervisors was fairly high, and ranged from 12 to 18 yr. of formal education. Most of the supervisors had a background in accounting or finance.

Administration Procedures and Instruments

A questionnaire was distributed and each participant was allowed about 1 wk. to complete it. The questionnaires were collected at group sessions during which research instruments for another study were being administered. All precautions to preserve anonymity were carefully observed.

Wernimont Job Satisfaction Scale.[2] Wernimont's (1966) scale, which was devised to measure present satisfaction according to a model constructed by Herzberg et al. (1959), was used in this study. For example, satisfaction with security is measured by the following set of statements:

—— I feel happy with the amount of security I have on my job. (5)
—— I feel somewhat happy with the amount of security I have on my job. (4)
—— I feel that the amount of security on my job is okay. (3)
—— I am not too happy with the amount of security I have on my job. (2)
—— I am not happy with the amount of security I have on my job. (1)

(The numbers in the parentheses following each item indicate the scale value of that item.)

The respondent is instructed to check two statements from each variable set which best describe his present satisfaction. The satisfaction score for each source is the sum of the two items selected. The higher the score, the greater the degree of satisfaction.

Similar sets of statements are used to measure motivator satisfaction (recognition, achievement, work itself, advancement, and responsibility) and hygiene satisfaction (department policy, technical competence–superior, interpersonal relations–superior, working conditions, interpersonal relations–peers, and, here, satisfaction with pay).

One additional item asks the respondent to indicate his present overall job satisfaction on a 9-point scale (9 being most satisfied).

Job Involvement (JI) Scale. The scale used to measure job involvement was initially developed by Lodahl and Kejner (1965). Originally, a 40-item scale was developed; this was then reduced to 20

[2] This scale is copyrighted by P. F. Wernimont and was used with his kind permission.

items "by considering the item-total correlations, the communality of an item, and the factorial clarity of the item [Lodahl & Kejner, 1965, p. 28]." Lodahl and Kejner established the split-half reliability and the concurrent validity of this 20-item scale by administering it to three occupational groups: nurses, engineers, and students. Construct validity was established by relating job-involvement score to Ghiselli Self-Description Index, Smith's Job Description Index, and the Ohio State Leader Behavior Description Questionnaire scores within the various groups of *Ss* employed (Lodahl & Kejner, 1965, p. 31).

In order to develop a shorter JI scale for use when time and space are limited, Lodahl and Kejner selected six items from the 20-item scale on the basis of a principal components analysis of the data from the engineering and nursing samples. The split-half correlation of this scale was reported to be .57 and its reliability was estimated to be .73. The correlation between this scale and the 20-item scale was reported to be .87. The shorter version of the JI scale was used in this study.

Of the six items in the short version of the JI, the first five are scored in a positive direction, from 1 (strongly agree) to 4 (strongly disagree). The last item is scored in reverse. A higher total score indicates lower involvement. Although this scoring appears to indicate job alienation rather than job involvement, it resulted from the method used in the construction of the original JI scale (Lodahl & Kejner, 1965, p. 27). At that time, items were put into a Likert-type format and assigned scores of 1 to indicate "strong agreement" and 4 to indicate "strong disagreement" with scores of 2 and 3 indicating intermediate positions. As the item selection process which led to the construction of the shorter 6-item scale was continued, the original scoring direction was maintained. The items comprising the JI scale are listed below:

1. The major satisfactions in my life come from my work.
2. The most important things that happen to me involve my work.
3. I'm really a perfectionist about my work.
4. I live, eat, and breathe my job.
5. I am very much involved personally in my work.
6. Most things in life are more important than work.

Method of Analysis

Responses to the items on the two instruments were scored, and correlations were computed between scores for each of the sources of job satisfaction and job involvement. Totals were also computed for the motivator and hygiene sources, respectively, by adding scores for the sources in each category. Correlations between these two totals and job-involvement score were also calculated. Finally, the correlations between the satisfaction totals and overall satisfaction, and between overall satisfaction and job involvement were also computed.

RESULTS

Table 1 shows the correlations of the five motivator variables and the total motivator score with the JI scale score. Table 2 shows the correlations of the seven hygiene variables and the hygiene totals with the JI scale score.

Table 1 shows that job involvement is related to satisfaction with recognition, achievement, and responsibility, although in each instance the percentage of total variance in job involvement accounted for is relatively small. The relationships between job involvement and satisfaction with work itself, and between job involvement and advancement, are not quite significant at the .05 level. These results confirm the hypothesis: Job involvement increases with increasing levels of satisfaction with motivator variables.

Table 2 shows that, as predicted, job involvement is unrelated to all but one of the hygiene variables: satisfaction with interpersonal relations with the superior. This finding lends further support to our hypothesis since it indicates that, in general, satisfaction with hygiene variables seems to be unrelated to job involvement.

Tables 1 and 2 also show that, as predicted, the motivator total satisfaction score correlates significantly (−.30) with job involvement. However, as expected, the correlation between the hygiene total satisfaction score and job involvement is not significant. Although the difference between these two correlations, and between the percentage of variance explained in each case, is not great it

TABLE 1

CORRELATIONS OF MOTIVATOR VARIABLES AND TOTAL WITH JOB INVOLVEMENT

Satisfaction variable	Correlation with job involvement
Recognition (a)	−.25*
Achievement (b)	−.21*
Work Itself (c)	−.19
Advancement (d)	−.19
Responsibility (e)	−.27**
Motivator total (Σ scores a+b+ . . . +e)	−.30**

Note.—N = 92; 4 *Ss* did not respond to the JI scale.
* $p < .05$, $df = 90$.
** $p < .01$, $df = 90$.

TABLE 2

CORRELATIONS OF HYGIENE VARIABLES AND TOTAL WITH JOB INVOLVEMENT

Satisfaction variable	Correlation with job involvement
Security (f)	.03
Salary (g)	−.13
Policy (h)	−.10
Technical Competence–Superior (i)	−.15
Interpersonal Relations–Superior (j)	−.22*
Working Conditions (k)	−.16
Interpersonal Relations–Peers (l)	.04
Hygiene total (Σ scores f+g+ . . . +l)	−.18

Note.—$N = 92$; 4 Ss did not respond to the JI scale.
* $p < .05$, $df = 90$.

does lend further support to the hypothesis of this study.

Table 3 presents the correlation of the motivator and hygiene totals with overall satisfaction and also indicates the percentage of total satisfaction variance accounted for.

Table 3 shows that the total motivator satisfaction and the total hygiene satisfaction scores correlate significantly with overall satisfaction scores. There is a relatively large difference in the percentage of total variance in satisfaction accounted for by each of these variables.

DISCUSSION AND CONCLUSION

Although the magnitude of the relationships was not large, increased job involvement did appear to be related to satisfaction with motivator variables. It therefore appears that the distinction between motivator and hygiene variables introduced by Herzberg et al. (1959) can be useful in predicting the job involvement of civil service supervisors. At the same time, two motivator variables, satisfaction with work itself and satisfaction with advancement, were not significantly related to job involvement. Advancement in the civil service, however, is based largely upon seniority and performance on competitive examinations. It therefore seems unlikely that this variable would necessarily function as a motivator.

TABLE 3

CORRELATIONS OF MOTIVATOR AND HYGIENE SATISFACTION TOTALS WITH OVERALL SATISFACTION

Total	Correlation with overall satisfaction	% of variance
Motivator (Σ scores on motivators)	.70*	49
Hygiene (Σ scores on hygienes)	.60*	36

Note.—$N = 96$.
* $p < .01$, $df = 90$.

In addition to the findings which supported the initial hypothesis, the results (see Table 3) concerning the relationship between overall job satisfaction and the total motivator and the total hygiene satisfactions seem to shed further light on the controversy surrounding the Herzberg et al. (1959) model of job satisfaction. The extent of this controversy is illustrated by two recent reviews of research relevant to the two-factor theory. In one such review, House and Wigdor (1967) reached the conclusion that the two-factor theory basically has no merit. In another review, Whitsett and Winslow (1967) came to the opposite conclusion and indicated that this approach to understanding job satisfaction has a good deal of merit. The majority of studies critical of the two-factor theory reached the consensus that both hygiene and motivator variables contribute linearly to overall job satisfaction. In the present study, this conclusion was also borne out, since total hygiene satisfaction and total motivator satisfaction both correlated significantly with overall satisfaction. However, total motivator satisfaction accounted for 49% of the variance in overall satisfaction, while total hygiene satisfaction accounted for only 36% of this variance. Although this study was not designed to test that hypothesis fully, the present findings seem to lend some support to the two-factor model of Herzberg et al. (1959). The major focus of this study was on the consequences of motivator and hygiene satisfactions for job involvement, and here, also, the results were supportive of the motivator-hygiene model.

REFERENCES

BASS, B. M. *Organizational psychology*. Boston: Allyn and Bacon, 1965.

HERZBERG, F., MAUSNER, B., & SNYDERMAN, B. B. *The motivation to work*. (2nd ed.) New York: Wiley, 1959.

HOUSE, R. J., & WIGDOR, L. A. Herzberg's dual-factor theory of job satisfaction and motivation: A review of the evidence and a criticism. *Personnel Psychology*, 1967, **20**, 369–389.

KATZ, D., & KAHN, R. L. *The social psychology of organizations*. New York: Wiley, 1966.

LODAHL, T. M., & KEJNER, M. The definition and measurement of job involvement. *Journal of Applied Psychology*, 1965, **49**, 24–33.

MARCH, J. G., & SIMON, H. A. *Organizations*. New York: Wiley, 1958.

WERNIMONT, P. F. Intrinsic and extrinsic factors in job satisfaction. *Journal of Applied Psychology*, 1966, **50**, 41–50.

WHITSETT, D. A., & WINSLOW, E. K. An analysis of studies critical of the motivator-hygiene theory. *Personnel Psychology*, 1967, **20**, 391–416.

(Received January 8, 1968)

Journal of Applied Psychology
1968, Vol. 52, No. 6, 474–479

TRAINING FOR VIGILANCE:

COMBINED CUEING AND KNOWLEDGE OF RESULTS [1]

EARL L. WIENER AND DENNIS A. ATTWOOD [2]

Department of Industrial Engineering, University of Miami

To test transfer of training in a visual monitoring task, 44 *S*s were given a 48-min. training session in 1 of 4 conditions forming a 2 × 2 factorial design. The 4 training conditions were knowledge of results (KR), cueing (also known as prompting), KR and cueing, and a control group receiving neither training aid. 7 days later all *S*s performed the same task with no training aids. Results showed that the KR-trained groups detected significantly more signals during transfer, but the cueing groups showed no significant differences. The group trained with the combination of KR and cueing did no better than the KR-only group. With respect to commissive errors (false alarms), there was no significant difference due to KR treatments, but the groups receiving cueing during training made significantly fewer during transfer. All 4 groups showed a decline in detection performance over time periods in the transfer session.

In an effort to elevate detection performance in vigilance tasks, several experimenters have turned to transfer-of-training techniques employing immediate feedback, or knowledge of results (KR), as a training aid. For a variety of monitoring tasks, it has been shown that operators trained with KR continue to outperform those trained without it (NKR) in later sessions when the KR has been removed (Adams & Humes, 1963; Hardesty, Trumbo, & Bevan, 1963; Mackworth, 1964; Wiener, 1963, 1967). Transfer of training in monitoring has been reviewed by Wiener (1967).

Recently another training aid has made a bid for attention: This method is known variously as guidance, prompting, or cueing. While KR is response contingent and occurs after a signal, missed signal, or false alarm, cueing takes place before the onset of the signal, alerting the operator that it is about to appear on his display. The leading proponent of cueing in vigilance tasks is Annett, who states that cueing is superior to KR as a training device. Annett and Clarkson (1964) found superior transfer in the cueing group compared to the KR group, but Annett and Paterson (1966) found no differences. Annett (1966, Experiment III) also reports cueing superior to KR in the transfer session; however, in that study the KR was not immediate feedback, but summary information provided every 5 min. In a recent review of five of their experiments, Annett and Paterson (1967) conclude that both KR and cueing increase detections, but that KR also tends to increase false reports. Finally, Colquhoun (1966) attempted to compare the transfer effects of various cueing and KR training schemes, but failed to find any differences between the experimental and control groups. Training with cueing and KR has been reviewed by Aiken and Lau (1967) and Lau (1966), who conclude that the benefits of cueing as a training device may have been underestimated, relative to KR, due to the prominence of the law of effect in learning theory.

This experiment sought to investigate the relative merits of cueing and KR in achieving successful transfer of training in visual monitoring, and particularly to examine whether any advantage could be obtained by combining the two. The experiment is essentially a 2 × 2 factorial design, consisting of the fourfold combinations of KR and NKR, cueing and no cueing.

[1] This investigation was supported in whole by United States Public Health Service Research Grant No. UI00014 from the National Center for Urban and Industrial Health. The authors gratefully acknowledge the assistance of Mary Dutcher and the staff of the Department of Laboratories of the Engineering School, University of Miami.

[2] Now at Defense Research Establishment, Toronto.

METHOD

Subjects

The *S*s were 44 undergraduate students in elementary psychology courses, 28 male and 16 female. They were recruited for the experiment with the offer of an hour's credit toward their experimental requirement or $1.50 for each of the two sessions. The *S*s were randomly assigned to groups with the exception that the male:female ratio was kept balanced. The groups are designated by their training conditions: KR (KR, no cueing), Q (cueing, no KR), KR + Q (both training aids), and N (neither training aid).

Apparatus

The task was an adaptation of one first used by Jenkins (1958), in which a voltmeter needle made 50 rightward deflections of 20°/min, the needle deflecting rapidly and returning to its resting position before the next deflection. The signal was a larger (30°) deflection. The meter faceplate was painted flat black and the needle white, and *S*s viewed the meter on small closed-circuit television monitors in a room separated from the apparatus. They responded to signals by pressing a hand-held cord switch. The deflections, both background and signal, and cueing were scheduled on punched paper tape, stepped through a reader at 50 characters per minute. The signal schedule consisted of 32 signals occurring randomly during a 48-min. vigil, with the restriction that eight signals occurred during each 12-min. period, and the minimum intersignal interval was .3 min. KR was provided automatically by a logic circuit which determined whether any response was a detection during the allowable time (2.5 sec.) or a commissive error. This information was displayed on a small lightbox located over the television sets. Detections resulted in a green light, commissive errors in red, and nondetections were indicated by an amber light which remained on for 5 sec. after the allowed response period. The box is similar to that shown in a previous study (Wiener, 1963, Figure 1). The KR box was removed from the booth during training for Group Q and Group N and during all transfer sessions. All *S*s wore earphones which played white noise from a Grason-Stadler Model 455C generator, to achieve auditory isolation.

Cueing was provided in the following manner: 3.6 sec. prior to a signal, the white noise was interrupted and a loud 1000-cps tone was presented in its place for 2.7 sec. The tone ended just prior to the signal deflection. The group receiving both cueing and KR (KR + Q) was simply provided with both the auditory cueing and the visual KR, and since they were separated by the signal and a 2.5-sec. response period, there was obviously no conflict or confusion between the two.

Procedure

The *S*s received initial instruction on the task and special instruction on KR and cueing, as appropriate to their group. The *S*s who did not receive cueing, KR, or both were not aware of its existence in other groups. Following the instruction and demonstration period, a standard 48-min. training session was run. The *S*s were reminded to return at the same time a week later, and on their return all *S*s performed another vigil with neither cueing nor KR. The *S*s in the groups that had received one (or both) of the training aids were simply told that it would not be present for this session. The signal schedule for the transfer session was the same as the previous session, but run in reverse, providing the same intersignal intervals with a different order. The data collected were detections and commissive errors (false alarms) by 12-min. periods.

RESULTS

The data collected consisted of the number of signals detected by each *S* during each of the four 12-min. periods, and commissive errors by periods. A commissive error was any response not preceded (within 2.5 sec.) by the presentation of a signal.

Detections

The detection data by groups and periods for the two sessions are shown in Figure 1 in terms of percentage of signals detected. In the transfer session all *S*s were run under conditions of no feedback and no cueing (identical to Group N in the training sessions), but are labeled according to the conditions under which they were trained a week earlier. The detection data for the two sessions were analyzed separately as two partially hierarchal analyses of variance, with *S*s nested in one of the four KR-Cueing treatments, but common to the four time periods during a session (Winer, 1962). The data were first subjected to Cochran's test for homogeneity of variance, and in neither analysis was the null hypothesis of equal variances rejected. The analysis of variance for the training session is shown in Table 1, and for the transfer session in Table 2.

Table 1 indicates that both training variables affected detection rate. Of course, one should place little importance on the significance of the cueing variable during the training session, as there is no reason why the cueing groups should have anything less than nearly perfect performance. The difference between Group Q and Group KR + Q was not statistically significant, according to a Duncan multiple-range test applied to the pair-wise

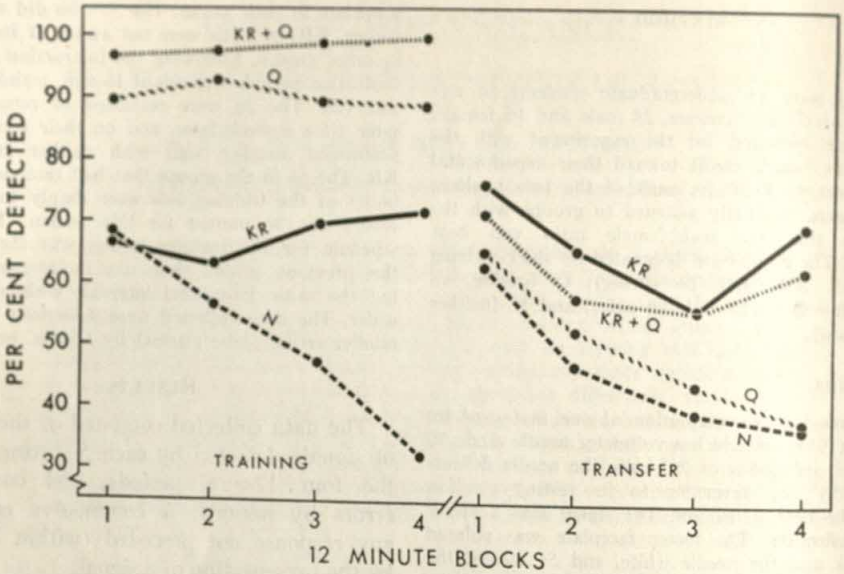


FIG. 1. Percentage of signals detected during training and transfer sessions as a function of training conditions and time periods.

differences between the four training conditions, but all other pair-wise differences were statistically significant. The fact that the KR main effect was statistically significant, as was the pair-wise difference between Group KR and Group N ($p < .05$), substantiates the familiar finding that KR produces superior monitoring performance. The significance of the KR × Periods interaction further supports previous contentions (Wiener, 1963) that KR groups often show no decline, and often even an increment in performance at the same time that the control group shows the familiar vigilance decrement.

The analysis of variance of the transfer session clearly indicates the transferability of differences produced by training with knowledge of results, while the differences produced by providing cueing in the training session, extreme as they were, failed to transfer after

TABLE 1

ANALYSIS OF VARIANCE OF SIGNALS DETECTED DURING TRAINING SESSION

Source	*df*	*MS*	*F*
Between *S*s	43		
KR[a] (A)	1	41.05	5.16*
Cueing (B)	1	330.01	41.45***
A × B	1	4.78	—
Error$_b$	40	7.96	
Within *S*s	132		
Periods (C)	3	2.07	1.35
A × C	3	6.57	4.28**
B × C	3	2.85	1.86
A × B × C	3	3.57	2.33
Error$_w$	120	1.53	
Total	175		

[a] Knowledge of results.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

TABLE 2

ANALYSIS OF VARIANCE OF SIGNALS DETECTED DURING TRANSFER SESSION

Source	*df*	*MS*	*F*
Between *S*s	43		
KR[a] (A)	1	72.55	6.71*
Cueing (B)	1	0.14	—
A × B	1	4.78	—
Error$_b$	40	10.82	
Within *S*s	132		
Periods (C)	3	21.52	9.70**
A × C	3	5.61	2.52
B × C	3	0.32	—
A × B × C	3	.26	—
Error$_w$	120	2.22	
Total	175		

[a] Knowledge of results.
* $p < .02$.
** $p < .001$.

a 7-day intervening period. Figure 1 indicates that cueing did not affect the performance of the group during the transfer session, and pair-wise range tests show no difference between Groups KR and KR + Q, or Groups Q and N. The significance of the period's main effect, coupled with the nonsignificance of interactions of the training variables with periods, points out that training techniques affect the overall performance level of the groups after transfer, but fail to prevent vigilance decrement.

Commissive Errors

Since commissive errors characteristically show an extreme positive skewness, the authors have adopted the practice of analyzing between-group differences of the number of *Ss* above or below the overall median number of commissive errors (see Wiener, 1967). Thus, the total number of commissive errors for each *S* was computed, and the median of these for the four groups combined was found to be 0.0 for the training session and 1.0 for the transfer session. The between-groups contrasts were then analyzed by forming a contingency table with groups as one dimension and the frequency of *Ss* above and below the median as the other dimension. The *Ss* making exactly the median number of commissive errors were evenly divided into the frequency categories. The data were then analyzed using the chi-square test for rom and column independence.

Since commissive error rates are meaningless while cueing is being provided to *Ss*,

TABLE 3

TOTAL COMMISSIVE ERRORS IN TRANSFER SESSION

Period	Group[a]				Total
	KR + Q	KR	Q	N	
1	4	6	1	10	21
2	3	10	7	7	27
3	2	15	4	6	27
4	1	10	6	11	28
Total	10	41	18	34	103

[a] Abbreviations: KR = knowledge of results, no cueing; Q = cueing, no KR; KR + Q = both training aids; N = neither training aid.

TABLE 4

NUMBER OF SUBJECTS MAKING LESS THAN, EQUAL TO, AND GREATER THAN THE OVERALL MEDIAN COMMISSIVE ERRORS

Errors	Group[a]			
	KR + Q	KR	Q	N
0	7	2	5	2
1	2	4	3	4
>1	2	5	3	5
Total	11	11	11	11

[a] Abbreviations: KR = knowledge of results, no cueing; Q = cueing, no KR; KR + Q = both training aids; N = neither training aid.

in the training session only the KR and N groups were tested. Of the 22 *Ss* receiving cueing, 19 made no commissive errors, and the remaining 3 made a total of 10. Considering only the KR and N groups, the overall median was three commissive errors, and the analysis of the contingency table revealed a nonsignificant difference ($\chi^2 = 2.2$, $df = 1$), indicating that KR did not affect the false-alarm rate.

The raw commissive error data for the transfer session are shown in terms of total commissive errors by groups and time periods in Table 3, and by numbers of *Ss* making less than, equal to, or greater than the median in Table 4. Median tests revealed that overall the four groups were not significantly different ($\chi^2 = 4.25$, $df = 3$), nor was the contrast of the two KR groups versus the two non-KR groups ($\chi^2 = 0.21$, $df = 1$). The analysis of the two cueing versus the two non-cueing groups indicated a significant difference ($\chi^2 = 3.84$, $df = 1$), with the combined cueing groups containing more *Ss* below the median number of false alarms.

DISCUSSION

The results of this study are further evidence that the most effective way to train a monitor for a monotonous watchkeeping job is to train him with immediate knowledge of results, and lend no support to the contentions of Annett and his collaborators that cueing, or prompting, is more effective. In fact, for this task, the authors were unable to find that cueing was effective at all, and

combining cueing with KR offered no improvement over the straight KR training method. To be sure, there are great differences between the auditory discrimination tasks used by Annett and his collaborators and the simple visual task employed here, and on a purely speculative basis the authors would agree with Lau (1966) that cueing may become more effective as the perceptual complexity of the task increases. In a very simple one-dimensional monitoring task, there is so little to be learned about the nature of the signal itself that the benefits of cueing, perhaps what Annett and Paterson (1967) view as providing the *S* with "an image or template against which further instances can be assessed," are minimized. They also regard KR not as a reinforcer, but as an inefficient form of retroactive cueing.

However, for a simple and monotonous monitoring task, the authors prefer to interpret the benefits due to KR in traditional law of effect terms, stressing more the motivational than the information component of KR. In very simple terms, *S* learns in the training session to maintain a higher level of alertness due to the reinforcing properties of KR, and he carries this tendency to the transfer session. The *S* may indeed obtain some retroactive cueing from the KR and learn something about the nature of the signal, but this should be viewed as a bonus dividend of KR and not its primary payoff. There is a strong possibility that *S* actually learns little or nothing as long as cueing is being provided, either because he may become "cue dependent" and pay little attention to the character of the signal itself, or because he may actually regard cueing as annoying. Some *S*s in this study volunteered unfavorable comments about cueing, describing it as irritating and "taking the challenge out" of the training session.

In summary, with the exception of the finding that the cueing groups produce significantly fewer commissive errors than the noncueing groups (as also reported by Annett & Paterson, 1967), the results of the present study, and of Colquhoun (1966), offer no support for cueing as a training aid in watchkeeping. Annett's strongest case is based not around detections, but the ability of cueing to reduce commissive errors (which the authors confirm), and the tendency of KR to encourage commissive errors (which the authors do not confirm, in this or in a number of other experiments of this type). One should recognize that the method in which *S*s were prompted in this experiment was quite arbitrary, and an entire array of questions on the most effective way of presenting the prompt remains to be answered. These questions involve such things as the length of the prompt, the sense modality, the duration between the end of the prompt and the onset of the signal, and a host of other hardware and software considerations. But based on the evidence to date, the authors are far more impressed by the potential of KR than any other training aids, as a practical solution to the many problems of man at watch. The authors will be eager to see if Aiken and Lau's prophecy, that in the future there will be as much research on cueing as there has already been on KR, will come to pass.

Finally, Swets and his collaborators (Swets, Harris, McElroy, & Rudloe, 1964; Swets, Millman, Fletcher, & Green, 1962) have examined the efficacy of computer-aided instruction in learning auditory identification, and Weisz and McElroy (1964) have found superior performance in visual identifications in a group combining cueing and KR. Noted recently is Mackie and Christensen's (1967) lament over the lack of practical application (to training) of the findings of experimental psychologists, and their call for "learning engineers" to tackle the problem. It appears that the monitoring task is ripe for the implementation of new technologies, and that the tools, ranging from hardware such as computer terminals and teaching machines, to software such as cueing, KR, signal schedules, and various tangible rewards, are readily at hand.

REFERENCES

ADAMS, J. A., & HUMES, J. M. Monitoring of complex visual displays: IV. Training for vigilance. *Human Factors,* 1963, **5**, 147–153.

AIKEN, E. G., & LAU, A. W. Response prompting and response confirmation: A review of recent literature. *Psychological Bulletin,* 1967, **68**, 330–341.

ANNETT, J. Training for perceptual skills. *Ergonomics,* 1966, **9**, 459–468.

ANNETT, J., & CLARKSON, J. K. The use of cueing in training tasks. NAVTRADEVCEN Technical Report 3143-1, 1964.

ANNETT, J., & PATERSON, L. The use of cueing in training tasks: Phase II. NAVTRADEVCEN Technical Report 4119-1, 1966.

ANNETT, J., & PATERSON, L. Training for auditory detection. *Acta Psychologica,* 1967, **27**, 420–426.

COLQUHOUN, W. P. Training for vigilance: A comparison of different techniques. *Human Factors,* 1966, **8**, 7–12.

HARDESTY, D., TRUMBO, D., & BEVAN, W. Influence of knowledge of results on performance in a monitoring task. *Perceptual and Motor Skills,* 1963, **16**, 629–634.

JENKINS, H. M. The effect of signal rate on performance in visual monitoring. *American Journal of Psychology,* 1958, **7**, 647–661.

LAU, A. W. A comparison of prompting versus feedback in verbal and perceptual learning. Tech. Bulletin STB 67-8, United States Naval Personnel Research Activity, San Diego, 1966.

MACKIE, R. R., & CHRISTENSEN, P. R. Translation and application of psychological research. Technical Report 716-1 (AD 650 309), Human Factors Research, Inc., Goleta, Calif., 1967.

MACKWORTH, J. F. The effect of true and false knowledge of results on the detectability of signals in a vigilance task. *Canadian Journal of Psychology,* 1964, **18**, 108–117.

SWETS, J. A., HARRIS, J. R., MCELROY, L. S., & RUDLOE, H. Further experiments on computer-aided learning of sound identification. NAVTRADEVCEN Technical Report 789-2, 1964.

SWETS, J. A., MILLMAN, S. H., FLETCHER, W. E., & GREEN, D. M. Learning to identify nonverbal sounds: An application of a computer as a teaching machine. *Journal of the Acoustical Society of America,* 1962, **34**, 928–935.

WEISZ, A. Z., & MCELROY, L. S. Response and feedback techniques for automated training of visual identification skills. NAVTRADEVCEN Technical Report 789-3, 1964.

WIENER, E. L. Knowledge of results and signal rate in monitoring: A transfer of training approach. *Journal of Applied Psychology,* 1963, **47**, 214–222.

WIENER, E. L. Transfer of training from one monitoring task to another. *Ergonomics,* 1967, **10**, 649–658.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received January 8, 1968)

Journal of Applied Psychology
1968, Vol. 52, No. 6, 480–483

INTERACTION OF TRAITS AND MOTIVATIONAL FACTORS IN THE DETERMINATION OF THE SUCCESS OF MANAGERS

EDWIN E. GHISELLI

University of California, Berkeley

Measures of the traits of intelligence, supervisory ability, and self-assurance, together with measures of the desires for job security, high financial reward, self-actualization, and power were obtained on 271 managers. In order to ascertain whether the motivational factors operated as moderators, on each of the motivational factors the managers were divided into those for whom the motivation was strong and those for whom it was weak, and separately for each group the relationship between scores on the 3 trait measures and job success was determined. It was found that in ½ of the instances the relationships between the trait measures and managerial success were significantly moderated by the measures of motivation thus indicating that there are interactive effects between traits and motivational factors in the determination of the success of managers.

Measures of various sorts of traits, both traits of ability and traits of personality, have been found to be related to the degree of success attained in management positions. Various sorts of motivational factors, too, have been shown to play a part in determining managers' performance. While it seems an obvious point, the fact of interaction between traits on the one hand and motivational factors on the other has not been given much attention. Even though it is possible to account reasonably well for variation in the job proficiency of managers in terms of the joint effects of traits and motivations in simple additive terms, because of the dynamic character of motivation it would appear likely that some model other than a linear combination of variables would be much more effective.

Unfortunately the term interaction is not easily defined, though obviously it implies that the joint effects of two or more variables are not of some simple additive sort. Certainly different sorts of methodologies reveal different kinds of interactive effects. The interactive effects disclosed by analysis of variance give one sort of result, and the multiplicative rather than the additive combination of scores on two or more variables provides still a different indication. Recently there has been interest in the interactive effects as shown by moderator variables. Each of the three approaches provides pertinent information, but at the same time it cannot be said that any one is more effective than any other in indicating interactions. Likely they all show different sorts of departures from simple linear combinations of variables.

A proposition that generally is taken as basic in psychological measurement holds that errors of measurement and errors of prediction are completely random. It follows from this proposition that there are no reliable differences among individuals in the extent of these errors, and hence any apparent individual differences in such errors necessarily must be unrelated to all other variables. However, it has become increasingly apparent that this state of affairs is by no means invariably the case, for in numerous instances it has been found that the degree of relationship between two variables, an indication of the magnitude of the errors of measurement and of prediction, is in fact a function of a third variable (Ghiselli, 1963a). That is, when individuals are sorted out into subgroups on the basis of the scores they earn on one variable, the degree of relationship between the scores they earn on two other variables varies systematically over the subgroups. A variable which operates in this fashion, interacting with two other variables in determining the degree of relationship between them for different individuals, is termed a moderator.

There has been relatively little attention devoted to the nature of the variables which moderate the relationship between pairs of other variables. Intuitively, at least, some psychological properties seem to be rather "static," whereas others give the impression of being "dynamic." Thus traits such as the capacity to solve problems in syllogistic reasoning or in arithmetic tend to be thought of as static traits, whereas motivational factors such as the desire to express one's own individuality or to dominate others are likely to be considered dynamic. Again intuitively, it seems probable that variables of the first sort would combine in simple additive ways, but that when variables of the latter sort are combined interactive effects might well be found. If this be the case, then it would be expected that measures of motivational factors would moderate the predictive power of measures of traits.

The traits of intelligence, supervisory ability, and self-assurance seem to stand up as having significance to performance in sundry managerial positions (Ghiselli, 1963b). The relationships between scores on measures of these traits and success in managerial positions are modest, but nonetheless seem to be consistently positive. It has also been found that three sorts of motivational factors, the desire for job security, for high financial reward, and for self-actualization, also play a part in determining the success of managers (Ghiselli, in press). A fourth motivational factor, the desire for power over others, apparently is unimportant in this respect. The purpose of the present investigation is to ascertain whether or not these four motivational factors moderate the relationships between the three trait measures on the one hand and the job success of managers on the other. That is, for those individuals who are strong in a given sort of motivation, is the relationship between scores on a trait measure and job success different from that for those individuals who are weak in that motivation? If differences of this sort do occur, then it can be said that there are interaction effects between traits and motivational factors in the determination of success in managerial positions.

METHOD

A forced-choice inventory was used which not only provides indexes of the traits of intelligence, supervisory ability, and self-assurance, but also of the motivational factors the desire for job security, for high financial reward, for power over others, and for self-actualization (Ghiselli, 1963b, in press). This inventory was administered to 271 persons holding middle management positions in a variety of different types of business and industrial organizations. For each manager a rating of his success in his own organizational setting was obtained. These ratings were dichotomous, each person was judged as being either a more or a less successful manager. Of the 271 managers, 55% were classified in the upper category and 45% in the lower one.

On each of the four motivational factors the managers were divided into the upper and the lower halves as nearly as could be achieved on the basis of

TABLE 1

COEFFICIENTS OF CORRELATION BETWEEN THE SCORES EARNED BY MANAGERS ON FIVE TESTS AND THEIR JOB SUCCESS FOR SUBGROUPS FOR WHOM EACH OF THE FOUR DIFFERENT MOTIVATIONAL FACTORS ARE STRONG OR WEAK

Test	Motivational factors							
	Job security		High financial reward		Power over others		Self-actualization	
	Strong	Weak	Strong	Weak	Strong	Weak	Strong	Weak
Intelligence	.11	.01	.26	.03	.05	.25	.17	−.01
Supervisory ability	.24	.58**	.41	.42	.45	.39	.23	.47*
Self-assurance	.48	.00**	−.02	.36**	.35	.12*	.06	.37**
No. cases	134	137	101	169	154	147	121	150

* Difference significant at $p > .05$.
** Difference significant at $p > .01$.

their scores on the scales measuring those factors. Then separately for each of these two subgroups, those in whom each given motivation was strong and those in whom it was weak, the coefficient of correlation between scores on each of the three trait measures and job success was determined.

RESULTS

The findings are shown in Table 1. Of the 12 differences between pairs of coefficients of correlation, 4 reach statistical significance at better than the .01 level of confidence and 2 at better than the .05 level. The desire for job security clearly has moderating effects on the relationships between success and the traits of supervisory ability and self-assurance. For those managers who are strongly motivated toward job security the relationship between supervisory ability and success is lower than for those in whom that motivation is weak. The opposite is the case for the trait of self-assurance. The desire for financial reward moderates the relationship between self-assurance and success, there being no relationship at all between self-assurance and success for those who are strongly motivated for such reward, and a fairly substantial relationship for those in whom such a motivation is weak. The desire for power over others moderates the relationship between success and self-assurance, the relationship being higher for those managers for whom the desire for power over others is strong. Finally, the desire for self-actualization moderates the relationships between managerial success on the one hand, and both supervisory ability and self-assurance on the other, the relationships being higher for those for whom the desires are strong.

DISCUSSION

All of the four types of motivational factors function as moderators. The differences between pairs of coefficients of correlation are somewhat larger for job security than for the other three factors, so possibly job security is the motivational factor most likely to produce interaction between traits and success. The desire for power seems to have least importance as a moderator since for it the differences between the pairs of coefficients appear to be smallest. Therefore different types of motivational factors differ in their effectiveness as moderators. This means that in the determination of the success of managers, some sorts of motivations interact with traits pertinent to performance in managerial positions to a greater extent than do others. Furthermore, some traits seem to be more susceptible to interactive effects than do others. None of the differences between the pairs of coefficients of correlations reaches acceptable levels of statistical significance in the case of the trait of intelligence, whereas all four motivational factors operate as moderators in the case of self-assurance.

In a sense, the desire for security and the desire for self-actualization are opposite sorts of motivations. Those individuals who seek safety in their employment situations would be expected to avoid acting in creative ways, thereby inviting attention to themselves, whereas it is likely that those who seek to express their individuality and to use their talents to the fullest would not be concerned with security. The results of the present investigation suggest that with managers who eschew either goal, job security or self-actualization, supervisory ability plays a much more significant part in determining success. Only when either of these two sorts of motivational factors is relatively weak does the individual's ability to direct the activities of others assume a prominent role in determining his success in administrative and executive positions.

The four motivational factors dealt with in the present investigation can be classified into two general sorts. The desires for security and for power might be thought of as selfish motivations which protect and gratify the ego, and the desire for high financial reward and for self-actualization perhaps could be conceived of as arising from individual strivings, the one being the reward expected for individual action and the other the expression of individuality. Looked at this way it then appears that when either the selfish motivations are strong or the individual ones are weak, then self-assurance assumes a paramount role as a determiner of managerial success. That is, the belief in one's self has a positive effect upon manager's performance

only when his goals are selfish and are not an expression of his individuality.

REFERENCES

GHISELLI, E. E. Moderating effects and differential relativity and validity. *Journal of Applied Psychology*, 1963, **47**, 81–86. (a)

GHISELLI, E. E. The validity of management traits in relation to occupational level. *Personnel Psychology*, 1963, **16**, 109–113. (b)

GHISELLI, E. E. Some motivational factors in the success of managers. *Personnel Psychology*, in press.

(Received January 19, 1968)

Journal of Applied Psychology
1968, Vol. 52, No. 6, 484-490

TASK SUCCESS, TASK POPULARITY, AND SELF-ESTEEM AS INFLUENCES ON TASK LIKING[1]

ABRAHAM K. KORMAN

New York University

Report of 3 experiments which support the general hypothesis that there is a positive relationship between task success and task satisfaction for persons who hold themselves in high self-esteem but not for those whose self-esteem is low. Implications for a general theory of job satisfaction as stemming from the moderator influences of personal and task variables are discussed.

Considerable interest in recent years has focused on the extent to which liking for a task or job is dependent on the extent to which one is successful in the given situation. Thus, while some argue that task success is important to job satisfaction but not necessary to job dissatisfaction (cf. Herzberg, Mausner, & Snyderman, 1959), others have claimed that such a distinction is too limited and that it is vital for both job satisfaction and dissatisfaction (cf. Dunnette, Campbell, & Hakel, 1967). Whichever theoretical approach one adopts, and the evidence tends to support the latter somewhat more strongly than the former (cf. Dunnette et al., 1967), it would seem apparent that increasing interest is being taken in achievement as a possible key determinant of the satisfactions one may derive from one's job, particularly since experimental evidence continues to be reported which supports the hypothesis quite well (Locke, 1965, 1966).

Despite such findings, however, some reason to question the proposed relationship in terms of its generality has been indicated. For example, in one of the studies (Locke, 1965) supporting the basic hypothesis it is pointed out:

> The author would not predict that there is any automatic relationship between task success and task satisfaction; rather the relationship should be highest in those cases where *S*s are trying hardest to succeed and lowest (or nonexistent) in situations where they are trying least hard to succeed. If situations could be devised where *S*s were not trying to succeed, the author would predict that the correlation between success and satisfaction would be lower than the correlation for the same task where the *S*s were trying to succeed [p. 385].

Since a number of writers have emphasized that the desire not to succeed and not to do well is particularly likely to occur both as a function of the personality characteristic of the individual and of role demands and expectations in lower levels of bureaucratic organizations (cf. Argyris, 1964; Korman, 1967), it would seem desirable that this possible limitation of the task-success–task-liking relationship be tested. Should such a limitation be established, this would have important implications for those wishing to develop an adequate theory of those conditions which facilitate satisfaction with one's employment situation.

The major purpose of the experiments to be reported in this paper was to test this possible limitation to the task-success–task-liking relationship within a self-esteem balance theoretical hypothesis to be described below. In addition, as the experiments proceeded, other hypotheses were developed which were deemed to be of relevance to the understanding of task satisfaction. These will be discussed at a later point in the paper.

Theoretical Framework

The theoretical framework used in this research stems from, but is not identical to, conceptions that have been offered by Vroom (1964) and Lecky (1945), in that it is based on a "self-consistency" hypothesis as to satisfaction, consistency leading to satisfaction, and inconsistency to dissatisfaction. The difference between the framework used here and their previous proposals lies mainly along three dimensions. First, it lays explicit heavy

[1] Portions of this paper were presented at the meeting of the Eastern Psychological Association, Washington, D. C., April 1968.

emphasis on self-evaluation and self-perception as part of the cognitive structure, and suggests different sources as to consequent variations in self-perceptions. Second, it incorporates within a common framework conceptions of the self-deriving from past reinforcement history as well as contemporary determinants of self-esteem (i.e., experimental manipulations and other interpersonal evaluations). Third, the basic theoretical assumption from which the derivation tested in this paper was taken attempts to predict variations on the three major dependent variables of industrial psychology: work performance, work satisfaction, and work choice, within a single framework.[2]

Thus, the major assumption of the framework is as follows:

All other things being equal, individuals will (*a*) choose, (*b*) adjust their performance to, and (*c*) find satisfying those behavioral roles which will maximize their sense of cognitive balance or consistency.

This being the case, then the following derivation can be made:

Individuals will tend to find most satisfying those job and task roles which are consistent with their self-cognitions. Thus, to the extent that an individual has a self-cognition of himself as a competent, need-satisfying individual, then, to that extent, he will find most satisfying those situations which are in balance with these self-perceptions.

It is assumed further that one's self-esteem or self-evaluation is the extent to which he sees himself as a competent, personal, need-fulfilling individual. Differences in such self-evaluation may arise, for our purposes, in three ways. First, one may conceive of a relatively chronic level of self-esteem, that is, a relatively persistent personality trait that occurs relatively consistently across various situations. Second, one may conceive of an individual's self-perceived competence concerning a particular task or job at hand. This may arise as a result of differential learning experiences or the specific characteristics of the moment. Finally, one's self-esteem is also a function of the expectations which others have of us. As Tannenbaum (1962) has argued, to the extent that others think that (*a*) we are competent and need-satisfying and able and (*b*) they exhibit such thoughts by the kinds of behavior which they exhibit toward us, to that extent our self-perceived competence concerning the task at hand is increased and so is, we would predict, task performance. The rationale for this is that such interpersonal evaluation provides a base of "reality" along which we may guide our behavior, as well as providing the "reality" which we learn from our previous experiences (Deutsch & Gerard, 1955).

From these considerations it was predicted in the experiments to be reported below that task success and task liking should be positively related for high self-esteem and unrelated for low self-esteem. Such a prediction was based on what seemed to be the reasonable assumption that success in a given situation is equivalent to adequacy in that situation and hence is in "balance" for high-self-esteem people, but not for low. In addition, as we have indicated, certain additional predictions were made in the last two experiments, with these to be described at that time.

EXPERIMENT I

Method

The tasks involved in this experiment were Subtests 1 and 2 of the Watson-Glaser Critical Thinking Appraisal, Form AM (Watson & Glaser, 1952), and 12 words from the medium anagrams lists of Tresselt and Mayzner (1966). (The subtests used are titled (1) Inference and (2) Recognition of Assumptions.) The *S*s were tested in groups of 15–20 and were presented the tasks as part of a normative study of college students' opinions about various ways of measuring creativity and intellectual ability. "Success" and "Failure" were determined by *E* informing *S*s of "pass-fail" points (based on scores from pretests) prior to asking them to make their ratings on a 5-point scale, but *after* they had completed the task.

As in previous research by the author (Korman, 1966, 1967), self-esteem was determined by scores on the Self-Assurance scale of the Ghiselli Self-Description Inventory,[3] with the cutoff for "high" and "low" being the median on the nationwide norms. Since evidence for the construct validity of this scale is contained mostly in an unpublished manual, Tables 1 and 2, consisting of a series of studies by the author of this paper and others, are presented as partial evidence for such validity. An examination of the relationships which are signifi-

[2] A. K. Korman, A Hypothesis Relating to Work Behavior. Unpublished manuscript, 1967.

[3] Edwin E. Ghiselli, Self-Description Inventory. Undated.

TABLE 1

RELATIONSHIPS BETWEEN GHISELLI SELF-ASSURANCE SCALE AND OTHER PSYCHOLOGICAL MEASURES

Sample	Measure	Highs		Lows		p
		N	M	N	M	
Engineering students	Gough Adjective Check List, Self-Confidence Scale	14	52.6	20	46.6	.05
Industrial foremen	Miner Sentence Completion Scale (a projective test of organizational power orientation)	10	4.70	12	.60	.05
Business students	Crites Need for Social Service (Likert-type scale)	35	5.77	36	7.97	.05
Business students	Crites Need for Job Freedom (Likert-type scale)	35	9.60	36	8.22	.10
Industrial foremen	Frequency with which parents supervised *S*s' jobs and tasks[a]	15	2.93	23	2.30	.05
Industrial foremen	Frequency with which *S*s argued with their parents during teens[b]	15	2.73	23	2.34	.05
Liberal arts students	Marlowe-Crowne Social Desirability Scale[c]	89				
Business students	Bass Self-Orientation (forced-choice tetrads)	20	27.10	15	24.06	.10

Note.—Division between highs and lows was based on the median of national norms. All tests of significance are two-tailed.
[a] Low score indicates greater frequency.
[b] High score indicates greater frequency.
[c] The author is indebted to Virginia Dunda and Charles Miller for these data; $r = .16$, *ns*.

cant and those which are *not*, as presented in the two tables, seems to offer support for the interpretations which are placed on this scale.

Subjects. The *S*s for Task 1 consisted of 35 high-self-esteem (HSE) *S*s and 36 low-self-esteem (LSE) *S*s, defined as stated above, while the *n*'s for Task 2 were 30 and 55, respectively. (Fifty-one *S*s were administered both tasks.) Finally, for Task 3, there were 25 HSEs and 42 LSEs. All *S*s were introductory psychology students fulfilling course requirements.

Results

Table 3 summarizes the findings of this experiment, providing complete support for the hypothesis. There is a significant relationship between task success and task liking for HSE individuals, but no such relationship exists for LSE individuals, with this holding true for all three tasks.

Discussion

The complete support for the hypothesis found in Table 3 appears to suggest in a fairly conclusive manner, at least for experimental laboratory tasks, that task achievement is important for task satisfaction, but only if one is used to task achievement and sees it as "proper" and "appropriate" for oneself.

However, the question now arises for the LSE individual as to what satisfies him, if task achievement does not? The task-liking and/or job-liking response occurs for him as well as for the HSE individual, but apparently the determinants are different. Experiment II was designed to provide a preliminary answer to this question with its major hypothesis being that task liking for the LSE individual is primarily a function of social influence processes. That is, it was predicted that, for LSE individuals, if others like the task they would, and if others did not they would not; the logic for this hypothesis being that we tend to use others

TABLE 2

RELATIONSHIPS BETWEEN GHISELLI SELF-ASSURANCE SCALE AND RANKING AND RATING VERSIONS OF THE STEIN SELF-DESCRIPTION INVENTORY

Need scale	Ranking M		p	Rating M		p
	Highs	Lows		Highs	Lows	
Abasement	16.5	16.2	*ns*	3.6	3.5	*ns*
Achievement	6.2	9.1	.05	2.2	2.6	.05
Affiliation	5.3	5.9	*ns*	2.2	2.1	*ns*
Aggression	13.5	15.0	*ns*	3.7	3.8	*ns*
Autonomy	12.2	8.3	.01	3.5	2.9	.01
Blamavoid	11.9	10.0	*ns*	3.1	2.9	*ns*
Counteraction	6.9	11.3	.01	2.5	2.9	.05
Defendence	13.2	13.5	*ns*	3.4	3.3	*ns*
Deference	11.8	12.0	*ns*	3.2	3.2	*ns*
Dominance	10.8	12.6	*ns*	2.9	3.0	*ns*
Exhibition	11.4	11.1	*ns*	3.3	3.2	*ns*
Harmavoid	13.5	13.0	*ns*	3.2	3.0	*ns*
Infavoid	13.2	11.7	*ns*	4.0	3.6	.05
Nurturence	8.2	5.5	.01	2.8	2.4	.05
Order	7.4	9.8	.05	2.3	2.4	*ns*
Play	10.9	8.3	*ns*	2.9	2.4	.05
Rejection	9.9	12.1	*ns*	3.3	3.5	*ns*
Sentinence	10.1	9.4	*ns*	2.7	2.6	*ns*
Sex	6.9	6.3	*ns*	2.0	2.0	*ns*
Succorance	10.1	8.9	*ns*	2.8	2.6	*ns*

Note.—Low mean indicates strong need; $N = 69$ college students. All significance tests are two-tailed. The author is indebted to Charles Miller for these data.

TABLE 3

LIKING FOR TASKS AS A FUNCTION OF TASK SUCCESS AND SELF-ESTEEM

Task	High self-esteem			t	Low self-esteem			t
	M	SD	N		M	SD	N	
Inferences				1.95*				.49
Succeed	3.86	1.2	22		3.20	1.2	25	
Failure	3.08	1.0	13		3.00	.65	11	
Recognition of assumptions				2.38**				.54
Succeed	4.07	.62	14		3.38	1.2	29	
Failure	3.31	1.1	16		3.32	.83	26	
Anagrams				5.07**				1.20
Succeed	4.21	.68	14		3.42	1.3	26	
Failure	2.18	1.2	11		3.00	.82	16	

* $p < .05$.
** $p < .01$.

as a guide as to how we should evaluate various phenomena in the environment (including ourselves) (cf. Deutsch & Gerard, 1955). It was also predicted that within each task there would be no relationship between task success and task liking. On the other hand, for the HSE individual, it was predicted that there would be such task-success–task-liking relationships for each task (i.e., high "others like" and high "others do not like"). The prediction for these latter individuals as to "social influence effects" between the two tasks was less clear, however. On the one hand, there seemed to be no reason to expect a lack of a "social influence" effect for HSE. Yet, it also appeared logical that the task-success–task-liking relationship might mitigate this effect somehow. For these reasons, no specific prediction was made as to how much "social influence" might affect the HSE individual liking for a task, except that it was thought that the effect would be somewhat less than for the LSE *S*s.

EXPERIMENT II

Method

The task involved in this experiment was the same anagrams task used in the first experiment, that is, 12 words from the anagrams list of Tresselt and Mayzner (1966). In addition, all other procedures were exactly the same with the sole exception that, in this case, *after* the task was completed and the individual was informed of his "success" or "failure," he was also told by *E* that most other students "did" like the task or "did not" like the task. After such information was presented, *S*s were then asked to rate their own liking for the task on a 10-point scale. As before, the measure of self-esteem used was the Ghiselli Self-Assurance scale.

Subjects. All *S*s were introductory psychology students fulfilling a course requirement. The breakdown by self-esteem scores resulted in 57 HSEs and 72 LSEs, with the division point being the median on the national norms.

Results

The analysis consisted of an "unweighted means" ANOVA (Winer, 1962), analyzing each self-esteem group separately, the predictions being that there would be a pass-fail effect for the HSE individual and possibly a "social popularity" effect, whereas only the latter would be true for the LSEs. The results are summarized in Table 4 and indicate complete support for the moderator influence of self-esteem on the task-success–task-liking relationship. However, support for the differential importance of social influence processes on liking for a task, when success on the task is controlled for, as a function of self-esteem, does not occur. In fact, the effect occurs more strongly for the HSE group than it does for the LSE, a result opposite to the prediction, but the effect is not significant for either.

EXPERIMENT III

Method

This third experiment constituted a replication of the second experiment with the following exceptions:

(*a*) The task used was the Inferences task from the Watson-Glaser Critical Thinking Appraisal;

(*b*) In addition to *E* stating that other students liked or disliked the task, an accomplice[4] was used to say that he liked or disliked it and to request of the *E* that he state his opinion. All three of these opinions were confirming of one another and thus, hopefully, provided a stronger "popularity" manipulation effect than in Experiment II (although a request for opinions after Experiment II similar to the one we will describe below for Experiment III indicated that the "popularity" manipulation was believed by all *S*s).

(*c*) Most importantly, self-perceived competence for the task (a component of self-esteem as we have defined it) was manipulated by *E* in the following manner. In Treatment A, *S*s were presented with the task and told that this was the first time this normative opinion study was being done at New York University. Prior to this time the test had been administered at Ivy League institutions which were highly selective in their admissions policies and it was from these individuals that we had our knowl-

[4] The author is indebted to James Hoon for his assistance in this experiment.

TABLE 4

LIKING FOR TASKS AS A FUNCTION OF TASK SUCCESS, SELF-ESTEEM, AND OTHERS' LIKING FOR THE TASK FOR EXPERIMENT II

Statistic	High self-esteem				Low self-esteem			
	Succeed		Fail		Succeed		Fail	
	Others like	Others do not like	Others like	Others do not like	Others like	Others do not like	Others like	Others do not like
M	7.13	5.83	4.93	3.58	5.95	5.38	6.27	5.00
SD	2.2	3.1	2.7	2.7	2.1	2.4	2.3	2.8
N	12	18	15	12	19	21	22	10

Note.—High self-esteem succeed–fail $F = 9.53$, $p < .01$; low self-esteem succeed–fail $F < 1$; high self-esteem others like–do not like $F = 3.65$; low self-esteem others like–do not like $F = 2.59$.

edge as to what average performance and liking levels were. Now we (the researchers) were moving into more "mass" institutions where students were of a much wider range in quality and where the average student was considerably below the Ivy League student in ability. Hence, the researchers did not expect that they would get the same levels of performance here.

In Treatment B, the same basic procedure was used except that *S*s were told that previously the researchers had been working on junior and community colleges and were now moving up to more "quality" institutions where they expected the level of performance to be higher. In both this treatment and in Treatment A, well-known examples were given of "high" and "low" quality institutions with which the *S*s at the institution were being classified in order to solidify further the *E*s' evaluation of their abilities.

The effectiveness of the general experimental procedure was checked by asking each *S* to write a paragraph indicating his reaction to the whole research situation in which he was participating. Of 179 *S*s, only 1 *S* indicated that he did not believe the normative purposes of the study, while only 4 did not believe the fact that the *E*s worked in the universities and/or colleges indicated previously. These *S*s were, of course, eliminated. Of the 175 remaining, every paragraph written indicated that they took seriously the stated purposes of the study as indicated by the *E*. In addition, their comments indicated that the "popularity" manipulation was accepted as well as in Experiment II.

Subjects. All *S*s were introductory psychology students fulfilling course requirements, with 94 being run in the "high" condition and 80 in the "low." These were further broken down, as in the second experiment, into high others-like-it and low others-like-it, as well as pass-fail, with these breakdowns being indicated in Table 5.

Results

The analysis performed here used the same procedures as in Experiment II, but with different results. There is an interaction effect for the high group indicating that achieve-

TABLE 5

LIKING FOR TASKS AS A FUNCTION OF TASK SUCCESS, SELF-ESTEEM, AND OTHERS' LIKING FOR THE TASK FOR EXPERIMENT III

Statistic	High self-esteem				Low self-esteem			
	Succeed		Fail		Succeed		Fail	
	Others like	Others do not like	Others like	Others do not like	Others like	Others do not like	Others like	Others do not like
M	7.53	4.16	5.82	4.58	6.96	3.80	6.19	4.35
SD	2.0	1.9	2.7	1.7	1.8	2.0	1.4	2.6
N	28	37	17	12	30	20	16	14

Note.—High self-esteem succeed–fail $F = 1.81$; low self-esteem succeed–fail $F = 2.27$; high self-esteem others like–do not like $F = 23.7$, $p < .01$; low self-esteem others like–do not like $F = 34.4$, $p < .01$.

ment has an effect on satisfaction only for the condition where others "like" the task, but that there is no effect when others do "not like" it. There is an others like/dislike the task effect for both groups, but there is no pass-fail effect in either case for the low group.

DISCUSSION

Taken in context, the results of these experiments are highly consistent in supporting the general hypotheses advanced in the introduction to this paper. It seems apparent that the meaning of satisfaction and liking varies as a function of the self-esteem variable and what induces liking for a phenomenon in one group will not induce liking in the other. We may consider then that the balance hypothesis advanced here has received support similar to that which it has received in other papers (cf. Korman, 1966), and that claims as to the great significance of task achievement for satisfaction are oversimplified at best. For some people, it is apparent, task achievement is not a source of satisfaction nor is task failure a particular source of dissatisfaction. To be more specific, task achievement to the LSE person is apparently irrelevant to his satisfactions. In addition, it also appears as though a task must meet at least some standards of "social desirability" before achievement on it will serve as any kind of satisfaction at all. People for whom task achievement is a source of satisfaction will not feel such satisfaction if the task is not worth achieving on.

What, then, does determine the satisfied response of the LSE individual, assuming his satisfied response is a lawful response? Our hypothesis that social influence processes may be involved in that perhaps the LSE person uses other individuals' response as a guide to his own response in a given situation was supported under the "strong" manipulation, but not the "weak." Since authority (E) influence was also involved as well as peer evaluation, we have little way of knowing whether it was the additive aspects of the influence in Experiment III which were important or its qualitatively different aspects. (The fact that the "popularity" manipulation, as measured by postexperimental opinions, was equally accepted in both cases only adds to the inconsistency of the findings.) Further testing of this hypothesis seems to be necessary, particularly when it might be moderated in its effects by the self-esteem variable, something which did not occur here. One possibility which might be investigated is that the effects of self-esteem might be curvilinear, rather than linear, as recently suggested by Cox and Bauer (1964).

What does seem to be clear, however, from these results, are the implications they suggest as to how to increase the feeling of satisfaction one gets from task achievement. Apparently what appears to be necessary is that one must first change the individual in order that he sees achievement as befitting him and being appropriate for him. It is only after such personal change has taken place that he will be satisfied with achievement on a desirable task and gain pleasant effect from it. How does one introduce this in the organizational environment? There are various ways of varying one's feelings of self-esteem and self-competence through interpersonal evaluation, varying control procedures and leadership behavior which seem to be quite common in the organizational environment, as Argyris (1964) and Korman (see Footnote 2), among others, have pointed out, and which would seem to be quite germane to this hypothesis. Hence, it would seem that a program of experimental research using these organizationally relevant variables would be of considerable value in further testing this hypothesis, particularly when task characteristics are also varied, a type of change already of some interest and concern in the modern organization.

REFERENCES

ARGYRIS, C. *Integrating the individual with the organization.* New York: Wiley, 1964.

COX, O., & BAUER, R. Self-confidence and persuasibility in women. *Public Opinion Quarterly,* 1964, 28, 453–456.

DEUTSCH, M., & GERARD, H. A study of normative and informational social influence upon individual judgment. *Journal of Abnormal and Social Psychology,* 1955, **51**, 629–636.

DUNNETTE, M. D., CAMPBELL, J. P., & HAKEL, M. Factors contributing to job satisfaction and job dissatisfaction in six occupational groups. *Organizational Behavior & Human Performance,* 1967, **2**, 143–174.

HERZBERG, F., MAUSNER, B., & SNYDERMAN, B. B. *The motivation to work.* New York: Wiley, 1959.

KORMAN, A. K. The self-esteem variable in vocational choice. *Journal of Applied Psychology,* 1966, **50**, 479–486.

KORMAN, A. K. Self-esteem as a moderator of the congruency between self-perceived aptitudes and vocational choice. *Journal of Applied Psychology,* 1967, **51**, 65–67.

LECKY, P. *Self-consistency: A theory of personality.* New York: Island Press, 1945.

LOCKE, E. A. The relationship of task success to task liking. *Journal of Applied Psychology,* 1965, **49**, 379–385.

LOCKE, E. A. The relationship of task success to task liking: A replication. *Psychological Reports,* 1966, **18**, 552–554.

TANNENBAUM, A. Control in organization: Individual adjustment and organizational performance. *Administrative Science Quarterly,* 1962, **7**, 236–257.

TRESSELT, M., & MAYZNER, M. S. Normative solution times for a sample of 134 solution words and 378 associated anagrams. *Psychonomic Monograph Supplements,* 1966, **1**, No. 15.

VROOM, V. *Work and motivation.* New York: Wiley, 1964.

WATSON, G., & GLASER, E. *Critical-Thinking Appraisal: Form AM.* New York: Harcourt, Brace & World, 1952.

WINER, B. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received January 23, 1968)

Manuscripts Accepted for Publication in the *Journal of Applied Psychology*

Personality, Problem-Solving Procedure, and Performance in Small Groups: Thomas J. Bouchard, Jr.*: Department of Psychology, University of California, Santa Barbara, California 93106.

Fatigue and Performance Variability among Typists: Leonard J. West*: Division of Teacher Education, The City University of New York, 33 West 42nd Street, New York City, New York 10036.

Convergent and Discriminant Validities between the Cornell Job Descriptive Index and a Measure of Goal Attainment: Martin G. Evans*: School of Business, University of Toronto, 119 St. George Street, Toronto 5, Ontario, Canada.

A Study of Some EPPS Variables as Factors of Academic Achievement: R. P. Bhatnagar*: Reader in Education, Regional College of Education, Ajmer, India.

Pattern Analysis of Biographical Predictors of Success as an Insurance Salesman: Robert Tanofsky, R. Ronald Shepps,* and Paul J. O'Neill: Marketing and Field Management, Metropolitan Life Insurance Company, One Madison Avenue, New York City, New York 10010.

Effect of Perceived Expertness upon Creativity of Members of Brainstorming Groups: Panayiota A. Collaros and Lynn R. Anderson*: Department of Psychology, Wayne State University, Detroit, Michigan 48202.

Predictive Power over 10 Years of Measured Social Service and Scientific Interests among College Women: Lenore W. Harmon*: 219 Mitchell Hall, University of Wisconsin, Milwaukee, Wisconsin 53201.

Validity of Estimates by Clerical Personnel of Job Time Proportions: Stephen J. Carroll, Jr.* and William H. Taylor, Jr.: Department of Business Administration, University of Maryland, College Park, Maryland 20740.

Architecture School Performance Predicted from ASAT, Intellective, and Nonintellective Measures: Clifford E. Lunneborg* and Patricia W. Lunneborg: Bureau of Testing, University of Washington, Seattle, Washington 98105.

Work Shift, Occupational Status, and the Perception of Job Prestige: Ronald H. Bohr* and Arnold B. Swertloff: Coordinator, Psychosocial Research, Rehab Center, Building S-9, Philadelphia State Hospital, Philadelphia, Pennsylvania 19114.

Task Structure, Work Structure, and Team Performance: James C. Naylor* and Terry L. Dickinson: Department of Psychology, Purdue University, Lafayette, Indiana 47907.

Beyond Parkinson's Law: III. The Effect of Protractive and Contractive Distractions on the Wasting of Time on Subsequent Tasks: David Landy, Kathleen McCue, and Elliot Aronson*: Department of Psychology, University of Texas, Austin, Texas 78712.

* Asterisk indicates author for whom address is supplied.

(Continued on page 496)

Journal of Applied Psychology
1968, Vol. 52, No. 6, 491–496

MULTIVARIATE RELATIONSHIPS BETWEEN A MEASURE OF VOCATIONAL INTERESTS AND A MEASURE OF VOCATIONAL NEEDS [1]

ROBERT M. THORNDIKE, DAVID J. WEISS, AND RENÉ V. DAWIS

University of Minnesota

The Minnesota Vocational Interest Inventory (a measure of the interests of nonprofessional men) and the Minnesota Importance Questionnaire (a measure of vocational needs) were administered to 505 male vocational rehabilitation clients. The group was split into 2 random samples of 246 and 259, and a canonical correlation analysis was developed on each group independently. Significant canonical correlations of .65, .56, and .52 were obtained for Group 1, and .62 and .54 for Group 2. The first 2 sets of weights for each group were cross-validated on the other group. The 1st set of canonical weights from each group yielded identical cross-validation coefficients of .38. Possible interpretations of the canonical variates and some theoretical and methodological considerations are discussed.

The purpose of the present study was to test the hypothesis that measured vocational interests and vocational needs reflect a common domain of psychometric behavior. Various attempts to relate interest measures to personality variables by zero-order or multiple-correlation methods have generally yielded low to moderate relationships. For example, Kohlan (1966) correlated the Strong Vocational Interest Blank (SVIB) and the Minnesota Importance Questionnaire (MIQ), a measure of vocational needs (Weiss, Dawis, Lofquist, & England, 1966). In predicting each individual SVIB scale from the 20 scales of the MIQ, Kohlan obtained multiple *R*s ranging from .28 to .62 with a median of .42. The low magnitude of relationship found is typical of that appearing generally in the literature, leading Thorndike, Weiss, and Dawis (1968) to argue for the use of canonical correlation for such multivariate problems. In reanalyzing Kohlan's data, Thorndike et al. (1968) obtained a maximum canonical correlation coefficient of .78 between the 45 SVIB occupational scales and the 20 MIQ scales. This coefficient accounted for 62% of the variance of the canonical variates. A replication study with these instruments yielded similar results.

The present study presents results on the canonical correlation of a different measure of vocational interests, the Minnesota Vocational Interest Inventory (MVII), with the same measure of vocational needs, the MIQ. A secondary objective of the present study was to determine the reliability of the obtained canonical relationships by subjecting the canonical regression weights to cross-validation.

METHOD

The *S*s for this study were 505 male applicants for services at the Minnesota Division of Vocational Rehabilitation. All *S*s completed both instruments in one testing session as part of a larger battery of tests given to assist in vocational counseling. The *S*s were administered the instruments in small groups throughout 1966 as part of the continuing research program of the Work Adjustment Project at the University of Minnesota.

After the data had been collected, the answer sheets were ordered according to a six-digit identification number and a table of random numbers was used to generate two groups of 246 and 259, respectively. Tests for differences of scale means and variances showed that the two groups did not differ in these respects. Possible effects of age, education, and marital status were also effectively minimized by the randomization procedure.

The data from the two samples were then subjected to independent analysis using the technique of canonical correlation, a correlational method which determines the maximum relationship between

[1] This study was supported, in part, by Research Grant RD-1613-G from the Vocational Rehabilitation Administration, United States Department of Health, Education and Welfare. Based on a paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1967.

two weighted linear composites (Cooley & Lohnes, 1962; Hotelling, 1936). The 21 MVII empirical scales (occupational keys) were treated as one set of variables and the 20 MIQ scales as the other. Due to the multiple and simultaneous determination of the regression weights in either set by the variables in the other set, the predictor-criterion distinction of the multiple correlation technique does not apply, and scores on either set of variables can be predicted from those of the other set. The computer program used (Cooley & Lohnes, 1962) calculates the canonical regression weights for the larger set of variables, then the weights for the smaller set of variables for each successive canonical correlation in turn. The canonical variates may be thought of as "factors" extracted from each set of variables such that the correlation between the "factor scores" (the canonical correlation) will be a maximum for each successive pair of variates. The magnitude of the correlation coefficient indicates the degree to which these "factor scores" (weighted linear composites of the original "predictor" and "criterion" variables) covary with each other or can be predicted from each other.

After the sets of canonical regression weights had been determined for each group, they were applied to the data for the other group in a double cross-validation design. Predicted total scores on each instrument were computed, based on the beta weights previously obtained from the other sample, and the product-moment correlation between these predicted total scores was obtained for each group.

Cross-validation of a canonical correlation is analogous to that of multiple correlation, except that the "actual criterion score" is the predicted total score of the "criterion" set.

RESULTS

The canonical correlations resulting from the development equations are shown in Table 1. For Group 1, three sets of weights yielded statistically significant canonical correlation coefficients of .65, .56, and .52, accounting for 43%, 31%, and 27% of the variance of the canonical variates, respectively. The data for Group 2 yielded two significant canonical correlation coefficients of .62 and .54 which accounted for 38% and 30% of the variance of these canonical variates.

Tables 2 and 3 list the scales scored on each instrument and the beta weights for the significant canonical correlations, rounded off to two decimal places. In using the weights for cross-validation, five decimal places were retained. It should be noted that the weights for the first canonical correlations of the two groups are more similar than the weights for any two successive canonical correlations for the same sample. This is to be expected due to the fact that succeeding pairs of canonical variates for a given sample must be orthogonal to all preceding pairs.

Table 2 shows that, for the first canonical correlation for Group 1, the variate defined by the MVII scales Milk Wagon Driver, Printer, Truck Driver, and Plumber was related to the variate defined by the MIQ Responsibility scale (positive weight) and by the MIQ Activity and Independence scales (negative weights). Table 3 shows that, for the first canonical correlation for Group 2, the variate defined by the Milk Wagon Driver, Stock Clerk, Printer, and Truck Mechanic scales of the MVII was related to the variate defined by the Responsibility and Supervision–

TABLE 1

CANONICAL CORRELATION, WILKS' LAMBDA, VALUE OF THE CHI-SQUARE STATISTIC, AND SIGNIFICANCE TESTS OF THE FIRST SEVERAL ROOTS OF THE INTERCORRELATION MATRICES BY GROUP

	Root	R_c[a]	Λ	χ^2	df	p[b]
Group 1[c]	1	.65	.066	610.0	420	<.001
	2	.56	.116	486.8	380	<.005
	3	.52	.168	403.1	342	<.05
	4	.49	.230	332.1	306	>.05
Group 2[d]	1	.62	.092	568.3	420	<.001
	2	.54	.149	455.0	380	<.05
	3	.47	.211	371.9	342	>.05

[a] The associ ated latent root, $\lambda_i = R_c^2$.
[b] Probability of error in rejecting the null hypothesis of no significant difference from zero canonical correlation.
[c] $N = 246$.
[d] $N = 259$.

TABLE 2

CANONICAL REGRESSION WEIGHTS FOR THE THREE SIGNIFICANT CANONICAL CORRELATIONS FROM GROUP 1

MVII weights				MIQ weights			
Scale	$R_{c1} = .65$	$R_{c2} = .56$	$R_{c3} = .52$	Scale	$R_{c1} = .65$	$R_{c2} = .56$	$R_{c3} = .52$
Baker	−.21	−.45	1.07	Ability Utilization	.05	.39	−.10
Food Service Manager	.19	.44	−.65	Achievement	−.16	−.22	−.01
Milk Wagon Driver	1.11	−.55	1.32	Activity	−.35	−.05	−.08
Retail Sales Clerk	.90	−.06	.64	Advancement	.28	−.10	.17
Stock Clerk	.23	−.38	−.21	Authority	.19	−.11	.34
Printer	1.11	−.32	−.32	Company Policies and Practices	.17	−.04	.15
Tab Machine Operator	.21	−.57	−.54	Compensation	.07	−.05	−.08
Warehouseman	−.66	−.37	1.34	Co-workers	.08	.16	.34
Hospital Attendant	.52	−.41	−.71	Creativity	.29	.39	.24
Pressman	−.01	−.45	1.55	Independence	−.46	.23	−.27
Carpenter	.17	−.34	−1.10	Moral Values	.06	.24	−.28
Painter	.14	−.40	−.27	Recognition	.14	.06	−.16
Plasterer	−.04	.18	.78	Responsibility	.35	−.53	−.52
Truck Driver	1.07	−.69	.30	Security	−.31	−.26	.29
Truck Mechanic	.20	−1.46	−.29	Social Service	.27	.07	.00
Industrial Education Teacher	.83	.54	.90	Social Status	.06	.17	−.02
Sheet Metal Worker	−.40	.13	1.60	Supervision—Human Relations	.08	.06	−.09
Plumber	.96	−.31	−.24	Supervision–Technical	−.02	−.23	−.08
Machinist	.32	−.69	.04	Variety	.20	.18	.29
Electrician	.28	−.40	−.39	Working Conditions	−.16	.10	.04
Radio-TV Repairman	.31	.46	.60				

Human-Relations scales (positive weights) and the Independence scale (negative weight) of the MIQ. Scales receiving large weights on the first pair of variates were found in common for each instrument and for both groups: Milk Wagon Driver and Printer (both received positive weights) for the MVII and Responsibility (positive) and Independence (negative) for the MIQ. Comparison of the weights for the second canonical correlation does not reveal the same similarity between groups.

By emphasizing those scales which show fairly reliable and large weights across the two samples an attempt can be made to interpret the canonical vectors. On the MVII the Printer and Milk Wagon Driver scales received consistently high weights. A vector defined in large part by these two scales is most highly correlated with a bipolar MIQ vector defined primarily by the Responsibility scale on the positive end and by the Independence scale on the negative end. Persons whose interest response patterns are similar to those of Printers and Milk Wagon Drivers also express preferences for job situations involving responsibility but lacking in independence (working alone). An independent validation of this relationship with appropriate occupational groups would be necessary before such an interpretation could be accepted as having been empirically demonstrated.

The weights which define the vectors of the second canonical correlation are so different for the two randomly developed groups that an interpretation of the above type seems unwarranted.

Table 4 contains the correlations obtained when the weights developed on Group 2 are applied to the data from Group 1 and vice versa. The first set of weights developed on Group 2 yielded a statistically significant correlation of .38 when applied to Group 1, while the second set of weights from Group 2 did not yield significant results. When the

TABLE 3

CANONICAL WEIGHTS FOR THE TWO SIGNIFICANT CANONICAL CORRELATIONS FROM GROUP 2

MVII weights			MIQ weights		
Scale	$R_{c1} = .62$	$R_{c2} = .54$	Scale	$R_{c1} = .62$	$R_{c2} = .54$
Baker	.07	−.11	Ability Utilization	.20	.01
Food Service Manager	−.72	.32	Achievement	.23	−.23
Milk Wagon Driver	1.07	.14	Activity	−.15	.44
Retail Sales Clerk	.68	.76	Advancement	.09	.10
Stock Clerk	1.01	−.69	Authority	.19	−.09
Printer	1.68	.22	Company Policies and Practices	.06	−.03
Tab Machine Operator	−.10	.29	Compensation	−.05	−.12
Warehouseman	−.22	.56	Co-workers	.17	−.15
Hospital Attendant	.42	.63	Creativity	−.03	−.21
Pressman	−.03	.84	Independence	−.46	−.23
Carpenter	.77	.52	Moral Values	−.03	.02
Painter	.43	.07	Recognition	.06	.33
Plasterer	.09	−.29	Responsibility	.56	−.04
Truck Driver	.68	.30	Security	−.14	.06
Truck Mechanic	1.29	−.67	Social Service	.02	.34
Industrial Education Teacher	.80	−.99	Social Status	−.04	−.21
Sheet Metal Worker	.25	−.34	Supervision—Human Relations	.37	.45
Plumber	.06	.79	Supervision—Technical	−.13	−.34
Machinist	.46	.68	Variety	.04	.12
Electrician	.56	1.75	Working Conditions	−.32	−.05
Radio-TV Repairman	.35	−1.43			

first two sets of weights developed on Group 1 were applied to Group 2, the first set yielded a correlation of .38 and the second a correlation of .28. Both of these correlations are statistically significant at $p < .01$.

DISCUSSION

The fact that significant correlations of .38 were found in both of the cross-validation canonical equations, which involved 41 weights each, is noteworthy. These results suggest a stable multivariate relationship between measured vocational interests and measured vocational needs.

TABLE 4

CROSS-VALIDATION CORRELATION COEFFICIENTS, F TESTS, AND SIGNIFICANCE LEVEL FOR THE FIRST TWO SETS OF WEIGHTS FROM EACH GROUP

Set	r	F	p[a]
Group 2 weights applied to Group 1			
First canonical weights	.38	41.7	<.001
Second canonical weights	−.05	.762	>.05
Group 1 weights applied to Group 2			
First canonical weights	.38	42.8	<.001
Second canonical weights	.28	21.5	<.001

[a] Probability of error in rejecting null hypothesis of no difference from zero correlation.

The authors attempted an interpretation of the first set of canonical weights, but only in the presence of cross-validation evidence which very strongly indicated that such an interpretation was justified. Other authors (e.g., Dunteman & Bailey, 1967) have interpreted canonical variates in cases where the evidence was less strong. Although those authors obtained two slightly larger correlations (.69 and .61) for the first two pairs of vectors, their sample was smaller and they reported neither a replication nor a cross-validation of their results. The present study shows that cross-validation or replication is required before any attempt at interpretation of canonical variates can be made.

The fact that a correlation of .54 became −.05 when the development weights were applied to a second random sample from the same population is suggestive of the amount of sample-specific variance that can be in-

volved in the calculation of canonical correlations. An interpretation of the second set of weights from Group 2 would have been meaningless for Group 1. It would therefore seem advisable to put little confidence in the interpretation of a pair of canonical variates developed from a single sample unsupported by additional evidence in the form of either a replication which produces very similar results or a cross-validation. Unfortunately, such interpretation is commonly included in the meager but growing number of research reports using canonical correlation techniques.

The results of this study lend support to an important trend in research methodology. Various authors, particularly Dunnette (1963), have argued that a single criterion measure is not adequate in the applied prediction situation because what is being predicted is very seldom a unidimensional attribute which can be adequately expressed as a single criterion variable. The use of a priori weights in a composite criterion is not very much better because the investigator seldom has a very accurate knowledge of what it is in the composite that he is in fact predicting. However, this less than desirable methodology has continued in common use because of the restrictions in the correlational techniques which have been used.

One of the weaknesses which is often mentioned in using the canonical correlation technique in an applied context is that the "criterion" weights developed to satisfy a mathematical least-squares solution may not reflect the values of the variables expected by the researcher. In this regard, it should be understood that canonical correlation analysis is primarily a research technique, the purpose of which is to show the investigator what his data imply. Canonical correlation analysis identifies the "criterion" variables which the "predictor" battery predicts and their relative importance from a linear-weighted composite point of view. If the "criterion" variables were designed to be used as part of a composite criterion, the canonical weights give the best regression weights for maximum prediction of that linear-weighted composite. Divergence of these canonical regression weights from the researcher's a priori weights should result in a reexamination and modification of the predictor battery so that the empirical and a priori weights become less discrepant, or in a modification of the researcher's a priori weighting system.

Although the method of canonical correlation has been known for over 30 years, the difficulties of computation have caused it to be generally disregarded until quite recently. This method, although still a tool for research rather than practical prediction, allows the applied psychologist to explore the variables of interest to him in the complex multivariate manner in which they occur.

In the context of the present substantative question, that of the relationship between measured vocational interests and measured vocational needs, canonical correlation has served to uncover a relationship hidden by other techniques. The present study, in conjunction with the previous study using the SVIB (Thorndike et al., 1968), suggests a moderate (cross-validated $R_c \simeq .38$) relationship between vocational needs and the measured interests of nonprofessional men, and probably a relationship of similar magnitude (un-cross-validated $R_c \simeq .75$) between measured vocational needs and the measured interests of professional men. Further, it has been shown that these relationships do not exclusively reflect the operation of chance factors in that cross-sample stability exists for the magnitude of the relationship and, in the present study, that the pattern of weights defining the relationship yields statistically significant results when applied to another sample from the same population. These data, therefore, suggest that vocational interests and vocational needs reflect the operation of similar underlying variables.

REFERENCES

COOLEY, W. W., & LOHNES, P. R. *Multivariate procedures for the behavioral sciences.* New York: Wiley, 1962.

DUNNETTE, M. D. A note on *the* criterion. *Journal of Applied Psychology,* 1963, **47**, 251–254.

DUNTEMAN, G. H., & BAILEY, J. P., JR. A canonical correlational analysis of the Strong Vocational Interest Blank and the Minnesota Multiphasic Personality Inventory for a female college population. *Educational and Psychological Measurement,* 1967, **27**, 631–642.

HOTELLING, H. Relations between two sets of variates. *Biometrika*, 1936, **28**, 321–377.

KOHLAN, R. G. Relationship between inventoried interests and inventoried needs in a college sample. Unpublished doctoral dissertation, University of Minnesota, 1966.

THORNDIKE, R. M., WEISS, D. J., & DAWIS, R. V. Canonical correlation of vocational interests and vocational needs. *Journal of Counseling Psychology*, 1968, **15**, 101–106.

WEISS, D. J., DAWIS, R. V., LOFQUIST, L. H., & ENGLAND, G. W. Instrumentation for the Theory of Work Adjustment. *Minnesota Studies in Vocational Rehabilitation*, 21, University of Minnesota, 1966.

(Received February 12, 1968)

(Continued from page 490)

Asymmetrical Transfer in Reading Texts Produced by Teleprinter and by Typewriter: E. C. Poulton*: Medical Research Council, Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.

Biographical Correlates of Artistic and Literary Creativity in Adolescent Girls: Anne Anastasi* and Charles E. Schaefer: Department of Psychology, Fordham University, Bronx, New York 10458.

Sources of Variation in Job and Life Satisfaction: The Role of Community and Job-Related Variables: Charles L. Hulin*: Department of Psychology, University of California, Berkeley, California 94720.

Vocational Choice and Professional Experience as Factors in Occupational Image: Edmond Marks* and Sam C. Webb: Office of Evaluation Studies, Georgia Institute of Technology, Atlanta, Georgia 30332.

Relation between Embedded Figures Test Performance and Simulator Behavior: Gerald V. Barrett, Carl L. Thornton,* and Patrick A. Cabe: D/419, Plt. H, Life Sciences Research Department, Goodyear Aerospace Corporation, 1210 Massillon Road, Akron, Ohio 44315.

Analyzing the Expert Judge: A Descriptive Study of A Stockbroker's Decision Processes: Paul Slovic: Oregon Research Institute, P. O. Box 3196, Eugene, Oregon 97403.

Color Coding Effects in Compatible and Noncompatible Display-Control Arrangements: Gary K. Poock*: Department of Operations Analysis, Naval Postgraduate School, Monterey, California 93940.

Effectiveness of SVIB Academic Interest Scales in Predicting College Achievement: Richard W. Johnson*: Counseling Center, University of Wisconsin, 415 Gilman Street, Madison, Wisconsin 53706.

Rated Acceptability of Mineral Taste in Water: II. Combinatorial Effects of Ions on Quality and Action Tendency Ratings: William H. Bruvold* and William R. Gaffey: University of California, School of Public Health, Earl Warren Hall, Berkeley, California 94720.

Personality Inventory Correlates of Creativity among Architects: Wallace B. Hall and Donald W. MacKinnon*: Director, Institute of Personality Assessment and Research, University of California, 2240 Piedmont Avenue, Berkeley, California 94720.

---

* Asterisk indicates author for whom address is supplied.

BUREAU OF EDUCATIONAL RESEARCH CALCUTTA

journal of applied psychology monograph

Vol. 52, No. 6, Part 2 December 1968

A SET OF BASIC INTEREST SCALES FOR THE STRONG VOCATIONAL INTEREST BLANK FOR MEN[1]

DAVID P. CAMPBELL, FRED H. BORGEN, SUZANNE H. EASTES, CHARLES B. JOHANSSON, AND ROBERT A. PETERSON

University of Minnesota

22 homogeneous content scales were developed for the Strong Vocational Interest Blank (SVIB) by clustering items with high intercorrelations. The resulting scales were named: Adventure, Agriculture, Art, Business Management, Law/Politics, Mathematics, Mechanical, Medical Service, Merchandising, Military Activities, Music, Nature, Office Practices, Public Speaking, Recreational Leadership, Religious Activities, Sales, Science, Social Service, Teaching, Technical Supervision, and Writing. Mean scores are presented on each of these scales for 113 occupational samples, and test-retest statistics are included for groups tested and retested over 2 wk., 30 days, 3 yr., 22 yr., and 30 yr. Similar information is presented for several student samples including a group who were tested as freshmen, sophomores, 5 yr., and 15 yr. after graduation. The main purpose of these scales is to provide some interpretation aids for the traditional occupational scales of the SVIB.

The regular scales now in use for the Strong Vocational Interest Blank (SVIB) were developed by comparing the responses of men in specific occupations with those of a men-in-general (MIG) group. Those items selected more often by the occupational sample serve as a scale for identifying the characteristic interests of the men in that occupation. These are usually termed "empirical" scales, and the scores provide an index of similarity between the individual's interests and the interests of men in the specified occupations. A substantial amount of research (see, e.g., Campbell, 1966; Darley & Hagenah, 1955; Strong, 1943, 1955) has shown that these scores are useful in psychological assessment projects and in counseling situations.

Problems with the Empirical SVIB Scales

The major advantage of these empirical scales is that they include all of the discriminating items for each occupation in one scale; this makes interpretation for that one occupation fairly easy. With these scores, a counselor can say something like, "You have interests similar to lawyers." However, the nature of empirical scales makes further psychological interpretation difficult. If the individual asks the obvious, "What does it mean to have interests similar to lawyers?" the counselor must fall back on what he has learned about lawyers from other sources. Although the related scales on the SVIB profile provide some general flavor, interpretation is still shallow except when done by well-trained sophisti-

[1] The statistics reported in this paper were expensive to accumulate, and many agencies provided assistance. The overall project was funded by the National Institutes of Health through Grant 1428-03, most of the computer time was furnished by the University Computer Center, University of Minnesota, and considerable clerical help was supplied by the Student Counseling Bureau, University of Minnesota.

© 1968 by the American Psychological Association, Inc.

TABLE 1

SAMPLE SVIB ITEM INTERCORRELATIONS

Item	% 500 Men In General				r
	Like	Indifferent	Dislike	Total	
Algebra					
Arithmetic					
Like	51	12	06	69	
Indifferent	03	10	06	19	.63
Dislike	01	01	10	12	
Total	55	23	22	100	
Geometry					
Chess					
Like	16	04	03	23	
Indifferent	24	12	06	42	.21
Dislike	15	11	09	35	
Total	55	27	18	100	
Actor					
Listening to a story vs. Telling a story					
Like	09	19	24	52	
Indifferent	06	10	16	32	−.19
Dislike	08	04	04	16	
Total	23	33	44	100	

cated counselors who have had considerable experience in working with the SVIB and who have studied the research literature.

A second major disadvantage is that there is no limit to the number of empirical scales; in theory one could be developed for each occupation—but there are over 20,000 occupations listed in the *Dictionary of Occupational Titles*. While one can, in only partial fantasy, visualize a computerized system where this information would be available for hundreds of occupations, this is still far beyond current capabilities. In recent practice, to compensate for this relatively narrow coverage of the occupational world, the counselor usually extrapolates from the available SVIB scales to other occupations not listed on the profile. For example, if a student inquires about geology, scores on the Engineer and Chemist scales are probably relevant. Some such extrapolation will be inevitable, no matter what the final method used, but it should be as easy and direct as possible.

A third disadvantage of these empirical scales is that they are difficult to work with in research studies. If an investigator wishes to study men who have survived in a specific occupational setting versus those who have not, he finds it cumbersome to compare these two groups on all of the 60 current empirical scales. While he can, and usually does, calculate the mean differences on each of the scales, the resulting statistics do not offer either a parsimonious nor easily interpretable method of understanding what distinguishes between the interests of these two samples.

The empirical success of the SVIB has created most of these problems. Because the scales are useful and do provide an easily interpretable score *for a specific occupation*, there has been considerable pressure to build more of them, further increasing the complexity of the profile. From all indications, this will continue as more occupations feel the need for self-study. This is good, for this detailed psychometric information on a wide variety of jobs is essential, but this prospect makes it even more imperative to develop a simpler way to summarize the results.

What both the counselor and researcher need is another system of scoring to supplement the occupational scales, a system containing relatively few scales, but scales which could be used to generalize beyond a single occupation.

Clark (1961), in his research with the Minnesota Vocational Interest Inventory, has shown that one way to do this is to work with measures that are homogeneous in content. Each of his homogeneous scales (or area scales as they are called on the profile) reflects interests in one type of activity or in closely related activities; thus, they are "pure" content scales.

Such a set of scales has been constructed for the men's form of the SVIB; this monograph is a report of their development and of their psychometric characteristics. Because this was an attempt to identify the important clusters of interests represented in the SVIB, these scales have been termed "basic interest" scales.

BASIC SCALE DEVELOPMENT

To build these scales, an item intercorrelation matrix was generated for the SVIB and,

TABLE 2

ADVENTURE CLUSTER

Item	211	14	231	12	334	86	277	337	321
211 Pursuing bandits in a sheriff's posse		24	26	37	23	44			
14 Airplane pilot	25		27	37	16	22			
231 Climbing along edge of precipice	23	12		25	19	10			
12 Auto racer	28	40	11		22	32			
334 Taking a chance	13	05	16	17		11			
86 Secret Service man	42	29	15	30	14				
277 Men who live dangerously[a]	21	22	20	25	16	27			
337 Thrilling, dangerous activities[a]	36	18	26	28	62	25	31		
321 Airline pilot[a]	18	51	13	23	12	18	22	30	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.

from this matrix, clusters of items with high intercorrelations were identified. The items in these clusters became the basic scales.

Set 1 and Set 2 of the Basic Scales

Actually, two sets of scales were constructed, the first using the 291 items common to all editions (1927, 1933, 1938, and 1966) of the SVIB booklet, the second using all 399 items of the 1966 revised form.

Although the first set will usually not be used in current projects either for counseling or research, three factors dictated its development: first, practically all of the SVIB criterion groups were studied with either the 1927, 1933, or 1938 booklets; second, many longitudinal studies have old data on an earlier booklet and current data on a later form; third, a more practical concern despite the efforts of the publisher and the scoring services, there inevitably are users who have not yet switched to the revised (1966) form. For all of these situations, a set of scales compatible with both early and recent booklets is needed.

The second set of basic scales is simply the first set expanded to include the items added to the SVIB in the 1966 revision. In this revision, a definite attempt was made to improve the coverage in areas relatively neglected in the original item pool such as art, religion, and music. As a result, scales for those areas are longer and more adequate in the second set. Unless confronted by one of the situations in the preceding paragraph, current users of the SVIB should use this second set.

Because most of the normative data presented here are from groups tested before 1966,

TABLE 3

AGRICULTURE CLUSTER

Item	37	102	342	345	76	191	230	162	214	150
37 Farmer		65	24	22	64	28	−34			
102 Agriculture	58		27	28	44	32	−18			
342 Outside work	21	27		40	30	26	−25			
345 Physical activity	20	19	40		18	25	−10			
76 Rancher	67	54	26	18		29	−30			
191 Handling horses	28	28	20	16	35		−16			
230 Living in the city	−24	−17	−24	−18	−21	−06				
162 Horseback riding[a]	17	20	13	11	27	71	−01			
214 Forest ranger[a]	34	37	23	18	37	24	−19	21		
150 Camping out[a]	18	20	20	13	18	23	−13	24	39	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] Item considered for inclusion in cluster, but discarded for various reasons. See text.

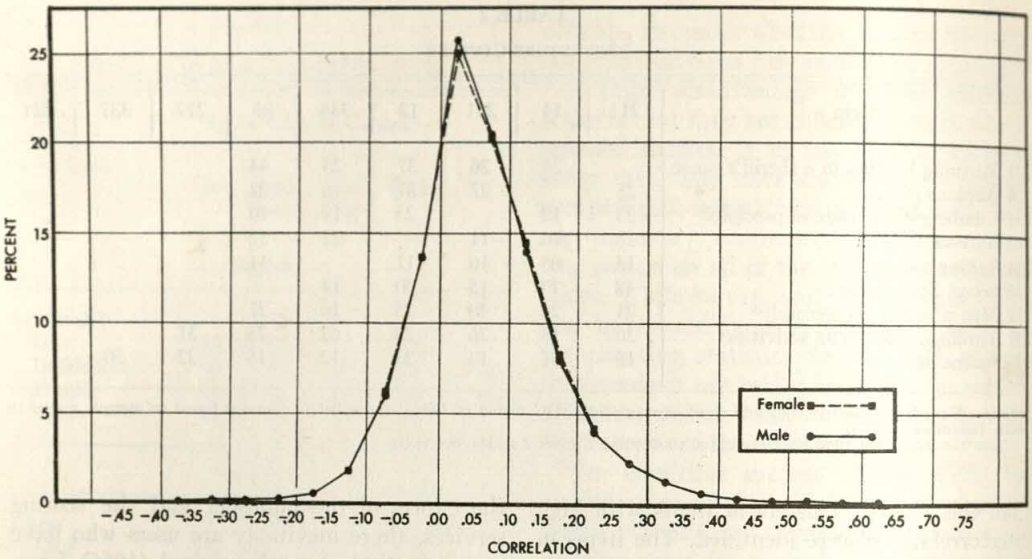


FIG. 1. Frequency distribution of item intercorrelations.

only the Set 1 scales are used unless specifically noted otherwise.

Eliminated Items

Two restrictions were applied to the individual items before scale construction began. First, to eliminate the highly popular and unpopular ones, no item was used if any of the responses, Like (L), Indifferent (I), or Dislike (D), had less than 15% response among a sample of MIG. Second, none of the items numbered 281–320 were included; those are the items grouped in tens where the respondent is to select the three he likes best, the three he likes least, and mark the remaining four indifferent. This format is troublesome as 1 or

TABLE 4

ART CLUSTER

Item	5	20	43	104	163	194	233	84	141	39	183	253	263	352
5 Artist		49	45	53	37	35	30	56						
20 Cartoonist	54		37	35	28	21	22	40						
43 Interior decorator	38	26		38	26	42	35	46						
104 Art	65	46	38		50	38	30	38						
163 Art galleries	49	36	34	50		32	36	37						
194 Decorating with flowers	33	21	42	35	36		39	33						
233 Looking at antique furniture	30	15	26	33	42	39		33						
84 Sculptor	67	48	37	57	51	30	31							
141 Sketching pictures of wild animals[a]	45	35	27	46	41	35	29	42						
39 Art museum director[a]	48	38	39	48	50	26	31	59	38					
183 Magazines about art and music[a]	44	29	29	49	60	34	35	48	37	46				
253 Artistic men[a]	50	31	39	49	40	33	30	45	34	39	40			
263 Prominent artists[a]	44	31	20	39	47	27	36	44	33	44	50	56		
352 Music and art events vs. athletic events[a]	35	21	19	33	42	15	18	38	26	41	50	22	34	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.

[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.

TABLE 5

BUSINESS MANAGEMENT CLUSTER

Item	33	35	58	65	81	206	226	347	196	42	85	96	177	278
33 Employment manager		44	27	44	36	30	29	−15	44	33	34			
35 Factory manager	47		49	41	34	32	32	−10	29	28	28			
58 Manufacturer	28	54		35	38	25	38	−11	27	18	22			
65 Office manager	57	48	34		50	33	36	−24	28	40	43			
81 Sales manager	47	42	37	42		36	42	−24	26	34	44			
206 Meeting and directing people	33	27	18	28	36		36	−17	39	31	26			
226 Developing business systems	30	34	34	39	41	29		−22	33	19	26			
347 Technical vs. supervisory responsibility	−30	−22	−17	−28	−37	−28	−26		−17	−18	−09			
196 Interviewing men for a job	50	31	21	40	39	45	33	−27		22	22			
42 Hotel manager	44	33	22	38	38	31	20	−30	30		35			
85 Manager, Chamber of Commerce	48	34	25	45	42	32	31	−22	30	38				
96 Travel bureau manager[a]	41	27	25	35	32	27	19	−17	25	43	39			
177 Business methods magazines[a]	34	37	37	33	45	24	48	−28	34	21	31	21		
278 Prominent businessmen[a]	28	31	33	30	35	29	30	−24	29	23	26	22	37	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.

2% of the respondents fill in this section incorrectly. Another 1% or so, aggravated by this forced-choice format, write some hostile comment in the margin and refuse to fill in any of the items. To solve these problems, and because these items are less efficient than the regular L-I-D items, this section will probably be removed in some future revision. In anticipation of that move, these items were not used in these basic scales. Only a few items were affected by these restrictions.

TABLE 6

LAW/POLITICS CLUSTER

Item	41	47	48	51	52	72	160	359	323	356	59
41 Governor of a state		51	42	31	35	51					
47 Judge	55		41	38	48	34					
48 Labor arbitrator	39	48		33	22	39					
51 Lawyer, criminal	45	50	35		42	36					
52 Lawyer, corporation	34	46	30	52		22					
72 Politician	62	48	44	43	34						
160 Electioneering for office[a]	50	33	35	34	28	65					
359 Dog trainer vs. parole officer[a]	−25	−23	−26	−23	−12	−24	−18				
323 Taxi driver vs. policeman[b]	−16	−14	−11	−23	−13	−11	−13	17			
356 Warden of a prison[b]	−01	−00	01	−06	03	−03	−09	10	10		
59 High school principal[b]	38	44	33	28	29	37	26	−24	−12	04	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.
[b] Item considered for inclusion in cluster, but discarded for various reasons. See text.

TABLE 7

MATHEMATICS CLUSTER

Item	101	103	107	115	120	128	148	122	24
101 Algebra		63	61	67	75	36			
103 Arithmetic	65		48	51	79	32			
107 Calculus	62	47		56	59	44			
115 Geometry	67	50	62		65	50			
120 Mathematics	76	77	63	66		42			
128 Physics	50	38	55	53	48				
148 Solve mechanical puzzles[a]	35	35	35	34	37	30			
122 Mechanical drawing[a]	26	31	31	35	36	35	41		
24 Civil engineer[a]	26	27	35	34	32	31	26	39	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.

[a] Item considered for inclusion in cluster, but discarded for various reasons. See text.

TABLE 8.—ITEM INTERCORRELATION MATRIX

Item	19	56	94	186	189	190	122	132	148	187	121	185	188
19 Carpenter		51	47	40	65	40	39	55	26	35	46	36	36
56 Machinist	47		66	47	44	62	41	55	34	55	40	48	48
94 Toolmaker	49	73		46	46	46	40	52	37	42	36	35	35
186 Repairing a clock	35	45	49		48	48	39	42	44	58	36	60	60
189 Cabinetmaking	57	43	49	48		46	47	57	33	41	51	40	47
190 Operating machinery	42	57	55	51	52		39	49	43	60	42	45	56
122 Mechanical drawing	32	41	42	38	44	40		62	41	35	61	38	39
132 Shop work	50	55	58	45	58	61	57		34	44	64	42	44
148 Solving mechanical puzzles	23	33	35	40	31	37	33	30		39	32	38	37
187 Adjusting a carburetor	33	52	55	66	45	62	33	52	39		37	63	69
121 Industrial arts	40	38	41	30	44	40	59	59	25	31		37	39
185 Making a radio or Hi Fi set	29	38	39	55	45	43	35	44	39	52	32		64
188 Repairing electrical wiring	32	45	47	62	53	60	39	50	37	67	30	53	
368 Have mechanical ingenuity	27	39	40	43	41	46	37	44	36	45	34	44	46
13 Auto mechanic	47	59	55	46	36	53	29	47	32	62	32	36	43
32 Electrical engineer	29	49	49	41	38	40	36	36	38	44	28	47	48
45 Inventor	24	30	32	30	33	34	19	26	31	30	21	36	32
55 Locomotive engineer	37	53	46	28	23	38	25	34	18	34	28	20	23
61 Mining superintendent	17	28	30	19	17	30	22	25	22	27	24	20	24
30 Draftsman	30	42	45	34	32	33	62	41	27	33	41	28	28
88 Shop foreman	33	52	54	29	29	42	33	42	25	40	34	22	31
24 Civil engineer	30	41	43	24	31	38	36	34	31	32	31	31	30
180 Popular mechanics magazines	32	45	45	45	39	50	43	51	38	49	42	42	46
98 Watchmaker	34	46	54	49	31	30	25	30	25	36	26	30	30
128 Physics	17	24	27	28	20	22	22	19	36	26	17	31	28
16 Designer, electronic equipment[a]	26	38	40	33	32	31	29	27	39	34	26	48	36
22 Electronics technician[a]	29	45	46	36	32	37	32	34	36	43	27	49	40
216 Looking at things in a hardware store[a]	33	34	36	30	37	39	25	39	23	32	29	28	36
34 Geologist[b]	30	24	26	20	27	18	15	19	21	19	19	25	21
87 Computer operator[b]	15	31	31	19	12	24	20	13	21	22	09	17	19
166 Science fiction magazines[b]	09	21	16	20	11	20	19	12	23	24	15	25	19

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in

[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.

[b] Item considered for inclusion in cluster, but discarded for various reasons. See text.

Item Intercorrelations

The next step in scale construction was to generate an item intercorrelation matrix. Using the responses from Strong's group of 500 MIG, a Pearson product-moment correlation was calculated between each pair of items by assigning the values +1, 0, and −1, respectively, to the Like, Indifferent, and Dislike responses, then treating one item as the x-axis, the other as the y-axis. Examples of three such correlations are shown in Table 1, including a high-positive, a low-positive, and a high-negative correlation. As there were almost no large negative correlations, except between forced-choice items where they were artificially created, −.19 qualifies here as "high."

The next step was to decide exactly how to build the scales. While the authors wanted to identify clusters of items with "high" intercorrelations, they initially had little idea of what to call high. To gain some knowledge here, a frequency distribution, using all of the correlations in the item intercorrelation matrix, was constructed; it is reproduced in Figure 1. To provide a replication, an analogous distribution is included for the item intercorrelations from the women's form of the SVIB. The two distributions are remarkably similar.

Several conclusions can be drawn from the distributions in Figure 1:

1. The item intercorrelations were generally low, mostly around 0, practically all between ±.20.

FOR MECHANICAL CLUSTER

368	13	32	45	55	61	30	88	24	180	98	128	16	22	216	34	87	166
33	49	28	26	37	32	38	42	32	24	37	30						
39	63	45	37	57	45	36	51	36	29	45	29						
36	54	39	35	40	51	40	57	33	18	55	32						
43	40	35	33	30	32	33	30	29	28	39	28						
41	36	29	35	30	31	40	39	34	31	39	34						
43	47	38	32	37	37	26	32	28	38	27	34						
39	28	31	27	25	30	55	31	39	32	22	35						
44	42	37	33	30	36	45	43	31	35	28	38						
36	27	36	39	23	30	33	30	26	34	29	30						
40	54	34	30	35	37	29	27	30	36	29	24						
42	35	24	29	22	24	40	30	25	33	20	27						
43	41	42	36	30	29	30	24	32	33	34	24						
41	38	39	29	31	27	31	29	25	34	28	24						
	26	28	41	25	22	28	26	20	32	22	34						
35		36	28	45	39	27	46	24	26	42	27						
33	40		41	40	41	34	42	56	27	27	37						
41	25	40		30	34	29	33	25	32	28	32						
13	42	29	17		37	18	39	23	24	35	25						
18	29	30	25	28		36	50	40	17	34	28						
27	33	40	17	31	22		38	37	24	34	27						
23	41	34	15	41	44	33		29	21	41	37						
30	34	57	34	29	36	39	30		14	24	31						
42	38	34	29	31	25	33	36	28		15	21						
23	42	33	23	38	24	31	31	24	27		24						
28	20	37	33	14	23	18	14	38	24	19							
34	37	65	45	22	30	34	26	50	30	33	41						
29	40	68	37	29	28	40	31	54	35	36	35	68					
26	26	24	19	25	21	19	24	**25**	43	18	18	20	23				
18	24	38	36	20	31	21	15	44	19	23	34	36	37	21			
04	24	32	13	24	20	31	35	22	17	24	18	30	36	12	17		
21	18	24	22	19	18	22	16	−08	33	21	17	26	26	14	15	22	

1960s. Decimals omitted.

TABLE 9

MEDICAL SERVICE CLUSTER

Item	29	67	69	93	130	136	192	256	264	49	356	64
29 Dentist		40	40	45	21	23	25					
67 Pharmacist	42		31	30	23	30	21					
69 Physician	38	32		78	31	23	36					
93 Surgeon	42	32	76		27	24	38					
130 Physiology	22	26	39	33		34	28					
136 Zoology	17	24	28	29	48		15					
192 Giving first aid assistance	49	46	47	46	41	38						
256 Physically sick people[a]	11	16	14	12	13	13	54					
264 Outstanding scientists[a]	11	15	20	18	22	28	07	07				
49 Laboratory technician[a]	25	52	16	23	17	25	40	06	25			
356 Superintendent of hospital[a]	12	10	22	20	20	13	28	04	07	06		
64 Psychologist[a]	14	15	25	20	35	19	35	05	14	14	10	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] Item considered for inclusion in cluster, but discarded for various reasons. See text.

2. The distributions were slightly positively skewed. This mild effect could be caused by any of several factors such as a slight response set or overrepresentation of some item-content areas.

3. There were not many high correlations. In the initial matrix, the highest were .79 between the items Mathematics and Arithmetic, and .78 between Surgeon and Physician, only 5% were above .25, and only 1% above .35. This strongly suggests that the items are relatively specific and that the respondents pay a

TABLE 10

MERCHANDISING CLUSTER

Item	80	99	81	219	18	42	85	220	65	92	68	346	96	177	218
80 Retailer		59	48	51	52	36	40	48	44	30					
99 Wholesaler	55		58	44	48	33	45	39	47	41					
81 Sales manager	52	48		47	48	34	44	44	50	38					
219 Buying merchandise for a store	53	49	50		54	32	31	61	32	26					
18 Buyer of merchandise	54	52	51	58		36	33	44	44	32					
42 Hotel manager	36	30	38	38	36		35	32	44	32					
85 Manager, Chamber of Commerce	34	36	42	36	36	38		30	47	41					
220 Displaying merchandise in a store	41	39	34	59	42	31	33		32	26					
65 Office manager	37	39	42	36	41	38	45	27		35					
92 Stockbroker	29	34	37	30	31	26	30	20	26						
68 Public relations man[a]	35	35	53	38	41	42	51	36	42	31					
346 Work in an import-export business[a]	28	30	36	31	35	33	24	28	20	21	37				
96 Travel bureau manager[a]	34	31	32	32	33	43	39	27	35	28	41	27			
177 Business methods magazines[a]	32	36	45	34	37	21	31	27	33	38	34	19	21		
218 Looking at things in a clothing store[b]	24	22	26	39	22	19	20	34	19	20	28	22	20	12	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.
[b] Item considered for inclusion in cluster, but discarded for various reasons. See text.

great deal of attention to the unique aspects of each item.

4. There were no large negative correlations. This indicates that, at the item level, there is little tendency for people who like one item to all dislike another. (This is peculiar, for large negative correlations certainly appear at the scale level, ranging up to −.88 between the Social Science Teacher and Engineer scales.) In constructing the basic scales, only a few items with negative correlations were included, even though a lower (absolute) value of correlation was accepted.

After the matrix had been calculated, and it

TABLE 11

MILITARY ACTIVITIES CLUSTER

Item	4	123	151	210	237
4 Military officer		49	42	59	
123 Military drill	52		66	68	
151 Drilling in a military company	53	81		61	
210 Drilling soldiers	59	68	74		
237 Military men[a]	52	44	43	47	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.

TABLE 12

MUSIC CLUSTER

Item	62	63	66	167	183	352	152	157	149	261
62 Musician		50	57	40						
63 Music teacher	56		52	27						
66 Orchestra conductor	58	51		34						
167 Symphony concerts	39	34	47							
183 Magazines about art and music[a]	43	41	46	54						
352 Music and art events vs. athletic events[a]	34	27	35	48	50					
152 Playing the piano[a]	49	37	42	30	30	22				
157 Jazz concerts[a]	26	20	28	25	30	08	29			
149 Religious music[a]	28	30	26	34	32	21	21	13		
261 Musical geniuses[a]	43	34	44	39	46	33	35	23	25	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.

TABLE 13

NATURE CLUSTER

Item	37	50	102	106	125	136	147	193	230	326	194	76	214
37 Farmer		40	65	33	35	26	34	40	−34	−14	19	64	
50 Landscape gardener	37		40	35	43	32	43	51	−17	−22	43	33	
102 Agriculture	58	42		42	44	34	39	42	−18	−08	29	44	
106 Botany	21	28	32		51	50	44	39	−09	−10	22	22	
125 Nature study	25	33	35	47		47	64	45	−08	−16	32	25	
136 Zoology	13	22	20	57	45		41	30	01	−05	21	24	
147 Bird watching	24	31	23	36	48	34		48	−17	−20	30	28	
193 Raising flowers and vegetables	33	44	40	28	30	16	24		−17	−32	51	24	
230 Living in a city	−24	−08	−17	00	−03	01	−06	−03		25	−01	−30	
326 Selling things house to house vs. gardening	−24	−34	−22	−19	−22	−12	−21	−26	13		−15	−11	
194 Decorating with flowers	15	41	23	25	27	22	29	37	02	−13		13	
76 Rancher	67	34	54	19	23	13	16	30	−21	−22	13		
214 Forest ranger[a]	34	30	37	22	37	21	26	26	−19	−24	17	37	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.

TABLE 14

OFFICE PRACTICES CLUSTER

Item	15	21	25	65	74	80	99	105	135	36	321	208	389
15 Bank teller		69	32	31	39	34	30	40	32				
21 Cashier in bank	81		25	35	36	37	37	43	27				
25 City or state employee	25	26		19	27	26	18	31	30				
65 Office manager	39	40	22		48	44	47	40	21				
74 Private secretary	35	33	22	31		38	34	35	32				
80 Retailer	32	30	13	37	19		59	32	25				
99 Wholesaler	30	29	09	39	19	55		33	23				
105 Bookkeeping	40	37	17	34	22	28	29		39				
135 Typewriting	20	17	13	20	17	14	15	28					
36 Income tax accountant[a]	43	40	17	34	24	23	55	19					
321 Airline pilot vs. airline ticket agent[a]	−18	−20	−15	−13	−08	−13	−04	−15	−07	−14			
208 Making statistical charts[b]	18	17	12	26	10	10	09	27	15	32	−04		
389 Pay attention to details[b]	−07	−03	−11	−10	−08	−06	−02	−14	−12	−13	14	−15	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.
[b] Item considered for inclusion in cluster, but discarded for various reasons. See text.

required 3 hr. with a CDC 1604 computer, various searching techniques were used, with the further aid of the computer when possible, to locate smaller matrices containing perhaps 5–25 items, all having high intercorrelations. Using Figure 1 as a guide, high was defined as above .30 though it was sometimes necessary to drop lower. The absolute minimum was set at .20 but even that was violated occasionally, especially when the correlations were negative.

Item Composition of the Basic Scales

Twenty-two basic scales were finally constructed; item intercorrelation matrices for each of them are listed in Tables 2–23. The items in the tables were sometimes abbreviated for easier presentation here.

In these matrices, two correlations are presented for each pair of items. Those above the diagonal were generated from Strong's 500 MIG, tested during the 1930s, and are the

TABLE 15

PUBLIC SPEAKING CLUSTER

Item	131	199	72	26	41	160	75	68	176	367
131 Public speaking		66	34	29	34					
199 Making a speech	67		39	31	29					
72 Politician	42	43		22	51					
26 Minister, priest, or rabbi	25	29	21		23					
41 Governor of a state	40	47	62	27						
160 Electioneering for office[a]	38	43	65	18	50					
75 Radio announcer[a]	34	32	32	24	30	29				
68 Public relations man[a]	38	28	39	21	34	35	39			
176 Telling jokes[b]	26	22	19	06	14	21	24	20		
367 Prefer working alone to working on committees[b]	−20	−18	−18	−08	−15	−17	−13	−18	−08	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.
[b] Item considered for inclusion in cluster, but discarded for various reasons. See text.

TABLE 16

RECREATIONAL LEADERSHIP CLUSTER

Item	7	70	79	127	143	171	12	350	280	60	352	174
7 Athletic director		55	44	44	32	41	26	25	34			
70 Playground director	55		40	27	13	27	22	20	20			
79 Reporter, sports page	51	43		26	31	49	26	24	23			
127 Physical education	62	42	40		26	27	16	26	31			
143 Boxing	32	24	25	32		33	20	20	19			
171 Sports pages in newspapers	39	29	48	38	20		20	14	31			
12 Auto racer	13	09	14	10	24	04		20	10			
350 Playing baseball	29	20	17	31	19	08	12		22			
280 Athletic men[a,b]	44	27	38	48	26	39	08	20				
60 Professional baseball player[b]	59	38	57	47	27	43	18	33	39			
352 Music and art events vs. athletic events[b]	−41	−16	−30	−31	−18	−49	−10	−07	−33	−37		
174 Skiing[c]	16	13	07	17	11	01	23	30	20	09	03	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] Inadvertently left out of Set 1 scales.
[b] New item from the 1966 booklet, added to develop the Set 2 scales. See text.
[c] Item considered for inclusion in cluster, but discarded for various reasons. See text.

correlations used here to cluster the items into scales. The correlations below the diagonal are based on a recently tested sample of MIG who filled in the 1966 booklet. (Many of them were actually tested in 1964 and 1965, using a prepublication version of this booklet.) This second group, which also included 500 men from a variety of occupations, provided a replication to recheck the homogeneity of the selected clusters. In general, the intercorrelations remained high in the second group, indicating that these clusters are stable over time and from one group to another.

The latter group also provided item intercorrelations for the new SVIB items added in the 1966 revision, and several items, those so noted, were added to the scales after reviewing these figures. The expansion of the scales with these new items constitutes Set 2 of the SVIB basic scales.

Considerations in Scale Construction

It is difficult to recapture exactly the methods used in searching out these submatrices; a combination of computer searching, art, hunches, and intuition was employed to decide which items should be grouped together, and it would be difficult to exactly replicate these techniques. To demonstrate some of these decisions, items at the bottom of some matrices are noted; these were considered for inclusion in the scale but were rejected for one reason or

TABLE 17

RELIGIOUS ACTIVITIES CLUSTER

Item	26	100	250	124	149	161	169	181
26 Minister, priest, or rabbi		52	46					
100 Worker in YMCA	30		36					
250 Religious people	35	30						
124 Bible history[a]	46	27	45					
149 Religious music[a]	39	30	42	46				
161 Going to church[a]	47	30	57	46	43			
169 Church young people's group[a]	35	45	52	46	46	59		
181 Reading the Bible[a]	39	26	51	61	52	56	51	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.

TABLE 18

SALES CLUSTER

Item	11	54	77	80	81	90	95	197	326	335	68	360
11 Auto salesman		43	49	33	32	42	49	35				
54 Life insurance salesman	43		49	33	36	41	42	41				
77 Real estate salesman	54	52		39	41	50	44	44				
80 Retailer	34	35	41		48	39	38	30				
81 Sales manager	43	41	48	52		47	48	54				
90 Specialty salesman	43	50	51	45	57		54	43				
95 Traveling salesman	43	44	45	38	52	56		51				
197 Interview prospects in selling	44	45	47	36	66	59	52					
326 Selling things house to house[a]	30	33	29	18	33	38	34	42				
335 Definite salary vs. commission on what is done[a]	−16	−15	−20	−07	−32	−28	−26	−35	−32			
68 Public relations man[a]	31	35	40	35	53	41	38	49	21	−17		
360 Appraise real estate[b]	25	21	33	17	27	21	18	28	20	−19	19	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.
[b] Item considered for inclusion in cluster, but discarded for various reasons. See text.

another, usually because of low intercorrelations but occasionally because of content.

To guide these decisions, the following two standards were closely followed: first, each decision was based on statistical evidence so that, for example, no two items were clustered together unless it was empirically defensible; second, the other major concern was with eventual interpretation. Thus, considerable attention was paid to item content. Occasionally, an item with relatively high correlations with a cluster of items was not included with them because the item simply seemed out of character. One such example was the item Sculptor in the Science cluster (Table 19). Although that item correlated fairly high with

TABLE 19

SCIENCE CLUSTER

Item	83	212	23	49	108	6	10	128	107	136	114	16	22	34	346	264	87	64
83 Scientific research worker		67	54	50	41	43	47	37	35	35	35							
212 Doing research work	67		42	39	36	29	41	41	34	32	27							
23 Chemist	57	44		52	63	35	39	33	36	38	30							
49 Laboratory technician	54	40	54		39	37	44	31	36	31	26							
108 Chemistry	38	30	57	37		25	25	50	38	36	30							
6 Astronomer	40	36	41	31	26		37	29	30	31	34							
10 Author of technical book	46	46	38	27	26	30		29	32	23	20							
128 Physics	41	36	45	32	50	34	33		44	30	32							
107 Calculus	33	29	34	22	40	30	30	55		19	19							
136 Zoology	33	28	31	25	28	32	30	27	21		41							
114 Geology	28	26	28	24	27	34	26	30	20	43								
16 Designer, electronic equipment[a]	50	41	55	46	37	38	39	41	31	20	22							
22 Electronics technician[a]	49	37	58	54	32	33	32	35	25	19	21	68						
34 Geologist[a]	44	37	44	39	32	47	32	34	24	40	61	36	37					
346 Business vs. research laboratory[a]	−50	−47	−35	−33	−27	−26	−31	−28	−23	−20	−12	−30	−31	−21				
264 Outstanding scientists[a]	31	27	30	25	26	24	32	28	17	28	24	23	19	18	−20			
87 Computer operator[b]	29	23	21	34	18	12	16	18	17	06	09	30	36	17	−16	14		
64 Psychologist[b]	22	15	12	14	08	20	20	12	08	19	12	10	06	17	−00	10	08	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.
[b] Item considered for inclusion in cluster, but discarded for various reasons. See text.

TABLE 20

SOCIAL SERVICE CLUSTER

Item	89	100	133	184	228	229	359	64		256
89 Social worker		59	49	40	26	32				
100 Worker in YMCA	46		34	28	30	35				
133 Sociology	47	23		41	22	33				
184 Social problem movies	49	24	47		14	20				
228 Contributing to charities	22	20	21	24		51				
229 Raising money for a charity	34	28	22	27	50					
359 Dog trainer vs. parole officer[a]	−29	−13	−26	−29	−15	−14				
64 Psychologist[b]	39	18	48	36	09	10	−04			
256 Physically sick people[b]	19	16	15	21	32	30	−10	05		

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.
[b] Item considered for inclusion in cluster, but discarded for various reasons. See text.

the science items such as Scientific Research Worker (.38), Chemist (.30), and Astronomer (.37), it was not included with them because it just did not seem to belong.

Another example is shown in Table 7 which has the item intercorrelations for the Mathematics scale. At the bottom of the table are several items which might have been included with this scale if the size of the intercorrelations had been the only consideration. However, they were not added because they would have clearly diluted the homogeneity of the Mathematics cluster. Such decisions were not made often, nor lightly, but all of the decisions were definitely not left in the hands of the computer—most of them, though, were.

Scanning the item intercorrelation clusters makes it clear that some item-content areas are better represented in the SVIB item pool than others. In the mechanical area, for example, many items were available, while in the Religious Activities area only three items were available for the Set 1 scales. Normally three items would not be considered sufficient to constitute a scale but because it was possible to generate a longer scale for the Set 2 scales such a scale was constructed.

While most of the possibilities for such scales in the SVIB item pool have been exhausted, no claims are made to have exhausted the domain of interests. When working with item intercorrelations, it becomes very apparent that a cluster of interrelated items will emerge if enough items are available concerning any one type of activity. But if an activity is not represented in the SVIB item pool, then

TABLE 21

TEACHING CLUSTER

Item	82	27	63	203	202	53	59	387	359
82 School teacher		68	48	38	36	40			
27 College professor	63		37	44	29	38			
63 Music teacher	34	26		18	21	39			
203 Teaching adults	48	44	23		45	18			
202 Teaching children	38	22	24	53		25			
53 Librarian	40	34	35	18	17				
59 High school principal[a]	57	46	29	32	28	31			
387 Have patience when teaching others[a]	23	22	13	34	26	06	18		
359 Dog trainer vs. parole officer[b]	−23	−23	−02	−20	−11	−04	−24	−17	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.
[b] Item considered for inclusion in cluster, but discarded for various reasons. See text.

TABLE 22

TECHNICAL SUPERVISION CLUSTER

Item	35	58	61	88	347
35 Factory manager		49	43	46	−10
58 Manufacturer	54		37	33	−11
61 Mining superintendent	40	38		50	10
88 Shop foreman	45	35	44		09
347 Technical vs. supervisory responsibility	−22	−17	01	−02	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.

no scale to measure interests in that area can be constructed. This is probably not a significant problem because the SVIB item pool does have considerable diversity, especially in occupational activities. However, certain omissions can be quickly identified by scanning analogous intercorrelation matrices which have been developed for the women's SVIB. For example, one scale available for the women's form but not the men's has been labeled "Homemaking" and contains the following items:

ITEM

Caterer
Cook
Housekeeper
Home Economics Teacher
Doing your own laundry work
Cooking
Sewing
Preparing dinner for guests
Trying new cooking recipes
Home Economics

Whether or not such a scale would be useful on the men's profile can only be an academic question here; because these items have never been included in the men's booklet, this scale is not available for use in comparing male occupational groups.

In constructing these scales, some difficult decisions arose over the issue of scale interrelationships, the most difficult being that of scale independence. Most psychometricians, and probably all factor analysts, would likely argue that these scales should be as independent as possible; certainly they would prefer that there be no overlapping items. While there might be some statistical advantages in such procedures, it is not at all clear that the resulting scales would be the most accurate reflection of the structure of interests, as some interest dimensions are related, and some SVIB items tap more than one dimension.

One such quandary appeared in the three scales: Sales, Merchandising, and Office Practices. The correlations between these scales are high: .78, .63, and .79, respectively, and there is probably only one underlying dimension—

TABLE 23

WRITING CLUSTER

Item	9	57	71	31	53	78	119	40	112	172	165	170	75
9 Author of novel		67	47	51	31	36	33	38	34	31			
57 Magazine writer	65		51	63	43	48	35	38	43	32			
71 Poet	52	50		41	41	33	31	30	32	51			
31 Editor	52	57	43		37	50	39	47	43	29			
53 Librarian	28	36	37	34		31	30	26	26	32			
78 Reporter, general	42	54	39	53	28		25	51	26	24			
119 Literature	51	42	48	43	30	40		26	54	40			
40 Foreign correspondent	47	51	38	51	25	54	36		29	18			
112 English composition	48	47	40	44	35	40	57	32		30			
172 Poetry	44	40	67	37	35	32	52	28	41				
165 Writing a one-act play[a]	58	64	55	50	37	46	43	45	46	44			
170 Biographies[a]	38	37	33	32	33	32	42	26	34	44	40		
75 Radio announcer[b]	23	34	22	36	15	39	22	34	19	14	33	16	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.

[a] New item from the 1966 booklet, added to develop the Set 2 scales. See text.

[b] Item considered for inclusion in cluster, but discarded for various reasons. See text.

business. Yet, there were three fairly distinct clusters of items and, as they clearly merit differing interpretations, they were retained as separate scales. Whenever such an arbitrary decision had to be made, the overwhelming consideration was to improve the final interpretability of the scales, not to satisfy intermediate psychometric requirements. The eventual impact of these decisions can be seen in Table 31 where the scale intercorrelations are reported; some of the correlations were quite high—in the .70s—and that is undesirable. Still, one cannot have it both ways; either you take the clusters as they stand, or you artificially eliminate overlapping items. The authors' preference was for the former.

"Personality" Scales

The scales which appeared first in the scale-construction stage were those that researchers in interest measurement have learned to expect: Sales, Science, Mechanical, Social Service, and so forth. But the authors were also interested in constructing scales that would resemble the more traditional personality dimensions such as aggressiveness or compulsivity.

Several attempts to build such scales from the SVIB item pool were singularly fruitless. Neither computer algorithms nor intuition located clusters of items that held together statistically. To illustrate the difficulties, intercorrelations for items selected to represent the dimensions of autonomy and compulsivity are shown in Tables 24 and 25. The level of homogeneity of these "clusters" falls far short of the minimal standards of statistical coherence which were followed in forming the occupational oriented basic scales, and these personality clusters were dropped. These failures might be attributed either to a lower cohesiveness of personality clusters, or to the inadequacies of the SVIB item pool in the personality domain.

Finally, to gain some experience with such a personality scale, the Adventure scale was forced into existence. Items with a derring-do flavor were separated out, their intercorrelations were scanned, and those with at least mild positive correlations, generally larger than .10, were clustered together. As will be seen later, the resulting scale worked reasonably well—a group of astronauts scored highest, 2.50 *SD*s above the lowest scoring occupations of school superintendents and mathematicians—but, still, the scale seems to have less occupational relevance than the other more homogeneous scales. For example, on the other scales, scores increase with age, especially with occupational experience, but scores on the Adventure scale *decrease* with age. In general, attempts to build homogeneous personality scales were unsuccessful.

Scale Reliability

Relatively few items are available in the SVIB for any single content area; thus, most of these scales are short and one of the first concerns here was with scale stability. Building short scales, no matter how pure, could conceivably result in scales with unacceptable reliabilities. To check this, several scales of varying lengths were constructed and their long-term reliabilities were compared. Six

TABLE 24

AUTONOMY CLUSTER

Item	299	335	336	351	338	367	392
299 Freedom in . . . doing work		−16	13	−06	−06		
335 Commission on what is done vs. salary[a]	−14		−25	−04	24		
336 Work for yourself	18	−29		−01	−31		
351 Amusement alone . . .[a]	−12	−00	07		07		
338 Work for self in small business[a]	−10	25	−48	05			
367 Prefer working alone	06	01	−00	−22	−01		
392 Dislike taking orders	08	−06	17	−09	−14	09	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.

[a] Item with negative weight.

TABLE 25

COMPULSIVITY CLUSTER

Item	105	223	227	334	341	344	380	362	366	373	385	389
105 Bookkeeping		20	26	−01	−09	−02	17					
223 Methodical work	21		14	−10	−21	−02	30					
227 Saving money	22	14		−24	−03	−08	15					
334 Playing safe*	−10	−14	−10		03	14	−06					
341 Work with many details*	−09	−27	−03	06		−08	−19					
344 Similarity in work*	−10	−10	−08	19	00		10					
380 Plan work in detail	12	28	18	−06	−18	−01						
362 Make decisions immediately	03	−11	−07	20	09	07	−10					
366 Keep detailed records of expenses	19	21	21	−15	−11	−13	29	−11				
373 Am always on time with my work	06	14	11	−11	−01	−09	25	07	13			
385 Am slow going and sure	−01	10	08	−21	01	−08	12	−29	06	−02		
389 Pay attention to details very little	−14	−26	−12	14	29	04	−37	17	−32	−13	−10	

Note.—Correlations above diagonal from Strong's 500 MIG tested in 1930s; those below diagonal based on sample tested in 1960s. Decimals omitted.
* Item with negative weight.

content areas were selected and two test-retest samples were used. The first samples included 102 members of an Army Reserve unit tested twice over a 30-day interval; the second included 191 Stanford University seniors tested first in 1927 and retested in 1949. Test-retest correlations for both of these samples for several experimental scales of differing lengths are reported in Table 26. Within each content area, the scales were formed from the "best" items available; thus, the 8-item Science scale has those items with the highest intercorrelations, the 16-item scale contains the initial 8, and the next 8, and so forth.

TABLE 26

RELIABILITIES FOR SOME EXPERIMENTAL BASIC SCALES

Content area	No. items	Corrected split-half reliability	Test-retest reliability	
			30 day	22 yr.
Science	8	.82	.86	.66
	16	.89	.91	.72
	24	.90	.88	.66
	32	.92	.91	.73
Mechanical	14		.93	.73
	21		.94	.70
	25		.94	.70
Sales	8		.89	.47
	16		.91	.50
Writing	4		.85	.60
	7		.83	.59
	9		.84	.61
	20		.88	.67
Public speaking	7		.87	.67
Mathematics	5		.88	.75

Note.—Median test-retest correlation for 53 regular SVIB scales: for 30 days, .91; for 22 yr., .67.

Several conclusions can be drawn from these reliability calculations:

1. The relationship between scale length and reliability was approximately the same, whether split-half or test-retest reliability coefficients were used. As the major concern with the SVIB is long-term stability, test-retest reliability is more important, and that statistic was used for the other comparisons.

2. In general, the expected relationship appeared between scale length and reliability—the longer scales were more reliable—but the relationship was by no means perfect. The shortest Mechanical scale (14 items) was more reliable over 22 yr. than the longest (25 items). The most reliable scale over the 22-yr. period was the 5-item Mathematics scale.

3. There was considerable difference between the content areas in the reliability of scales of equal length. Thus, the 8-item Sales scale, 7-item Writing scale, 7-item Public Speaking scale, and 5-item Mathematics scale had 22-yr. test-retest reliabilities ranging over a con-

siderable span: .47, .59, .67, and .75, respectively. This range could be caused either by differing degrees of homogeneity within the scales, the nature of the interest clusters involved, the composition of the test-retest samples, or all three factors.

4. In general, these reliabilities were roughly equivalent to those of the regular SVIB occupational scales. The loss due to scale shortness was apparently offset by the gain attributable to scale homogeneity. (The regular occupational scales have roughly 60–75 items.)

The general conclusion reached from this series of studies was that short scales, if their content is homogeneous, can be as reliable as long scales. Thus, the conflicting goals of purity of content, measured by item intercorrelations, and longest possible scale length were stressed in the scale construction.

Further information on the stability of these scales is reported in a later section.

Item Weights

Once the items had been selected for each scale, scoring weights were established by assigning the weight of +1 to the Like response, and −1 to the Dislike response. In the few cases where the item intercorrelations were negative, these weights were reversed.

Cranny's Work in Factor Scales

In a closely parallel but independent project, Cranny (1967) has factor analyzed the items of the SVIB; he described his project as follows:

> The responses of 500 men in general to each of the 288 items were intercorrelated and the items arranged in 15 clusters on the basis of their intercorrelations. Thurstone's (1947) multiple group method of factor analysis was followed. One factor was extracted from each cluster and the angular cosines among the 15 oblique factors obtained. Four factors were extracted from the 15 by 15 matrix of angular cosines using Lawley's (1940) maximum-likelihood procedure. The residuals were reduced to less than .00005.
>
> The four factors were rotated obliquely and Wherry's (1959) hierarchical factor solution was applied to yield four sub-general and 15 group factors. The factor loadings on the 19 factors of each of the 288 items were computed and a 288 by 288 residual table was obtained.
>
> The items with loadings on each factor are listed and descriptive labels are suggested for each factor. One of the group factors was not sufficiently defined to permit interpretation [p. 1].

TABLE 27

LISTING OF CRANNY'S VERSUS SVIB SCALES

Cranny scales	SVIB scales
Art	Art
Social Welfare and Service	Social Service
Public Contact	Public Speaking
Writer	Writing
Military	Military
Farming	Agriculture
Natural and Social Science	Nature Science
Mechanical	Mechanical
Undefined	
Medical	Medical Service
Mathematics	Mathematics
Business Management	Business Management
Sales	Sales
Musician	Music
Law	Law/Politics

Note.—Scales with no Cranny counterpart: Technical Supervision, Office Practices, Merchandising, Recreational Leadership, Adventure, Religious Activities, Teaching.

There is considerable agreement between Cranny's factor scales and the scales derived in the current project, which is interesting because the scale-construction techniques differed somewhat. Cranny's was essentially a statistical approach; in contrast, the methods of this study were much more intuitive.

A comparison of the two sets of scales appears in Table 27 where the scales are matched up by name. This comparison exaggerates the similarity slightly as some scales with the same names have slightly differing item content. Still, the resemblance is considerable and reassuring.

Many of the differences were due to greater concern in this present study for eventual interpretability. The treatment of the business area is one example; Cranny has two scales: Business Management and Sales, while this set contains five, covering approximately the same item content: Business Management, Sales, Technical Supervision, Merchandising, and Office Practices. There is no simple way to determine which approach is better. Cranny's scales are probably statistically more defensible; the present authors' scales are probably easier to interpret—so the choice between them is largely a matter of taste.

Cranny's report, which is an impressive amount of work for a doctoral dissertation,

contains no normative or other psychometric data on his scales.

Evaluation of the Basic Scales

In developing these scales, both statistical precision and common sense were used. The success of these techniques must be determined, not by studying the scale-building techniques themselves but by the efficacy of the resulting scales. To this end, a great deal of information follows on the reliability, validity, and other characteristics of these Basic Scales. The results seem meaningful and, after studying these data, users of the SVIB should have more confidence in their interpretations, and researchers should be able to home in more precisely on the unanswered questions.

One result of this work is that now something more is known about the organization of interests. The clusters of items were not forced; for the most part they fell out of the intercorrelation matrix relatively cleanly and their makeup reflects something of the basic dimensions underlying the items. In this sense, some clusters that did not appear are noteworthy. The failure of personality clusters to appear has already been noted. Another interesting omission was anything resembling a "prestige" cluster—it might have contained items such as corporation president, governor, and perhaps college professor. This might indicate that status per se is not an important determinant of item choice.

The clusters that were formed merit more extensive study as they have some psychological integrity, so to speak. More is said about this in the section on construct validity. One observation can already be made; that is, item clustering tends to follow vertical, not horizontal, divisions in the occupational world. Thus, the Science scale includes high and low laboratory positions; the Office Practices scale includes high and low office positions. This strongly suggests that the unifying feature is the activity involved, not the more tangential aspects of the work.

NORMS FOR THE BASIC SCALES

Once the basic scales were developed, norms had to be developed for interpretation; simply calculating the number of Likes that an individual marks in each scale would be of little value because the areas differ in popularity. To make the scores meaningful, they must be standardized in some manner that will permit ready comparisons between individuals and between scales. One common way to do this—and the method that has been employed here—is to convert the raw scores into standard scores with a known mean and standard deviation, usually 50 and 10, respectively. To do this, it is necessary to score some reference sample on each scale, then use the resulting raw score mean and standard deviation in a raw-score-to-standard-score conversion formula.

The question as to which reference sample should be used is difficult to answer. The regular occupational scales are normed against whatever occupation they are based on but that is not possible here as these scales are not specific to any one occupation. For these basic scales, there are various arguments for using a more general sample such as college freshmen, or high school seniors, or college-bound high school seniors, or perhaps a random sample of young adults, or even a stratified sample of the entire male population. Each would have its peculiar advantages and disadvantages.

Description of Norm Sample

The selection of a norming sample was again a result of concern with eventual interpretability. The men in this sample have been tested twice, once as teenagers and once as adults. They were first tested by Strong in 1930 when they were 16 yr. old, in his cross-sectional study of the relationship between age and interests (Strong, 1931). They have been retested in 1966–67 in a longitudinal study of the relationship between teenage interests and

TABLE 28

EDUCATIONAL LEVEL OF NORM GROUP

Educational level	*n*	%
Less than high school graduate	94	14
High school graduate	139	22
Some college	170	26
BA, BS	149	23
MA, MS	50	8
MD, DDS, LLB	26	4
PhD, EdD	19	3
Total	647	100

adult occupations. (When retested, their mean age was 51.8, with a standard deviation of 1.50. To describe them as "16-yr.-olds" when first tested and "52-yr.-olds" when retested provides brief and essentially correct labels.) For this reason, they make an especially good norm group as norms can be established, using the same people, for both teenagers and adults. This advantage of being able to compare directly 16-yr.-olds with 52-yr.-olds, using the same sample, is offset slightly by the disadvantage that the 16-yr.-olds were drawn from the 1930 population, not the current generation, but this is a small price to pay for having directly comparable groups.

A more precise description of the sample is not possible, for Strong never wrote down his specific 1930 sampling methods. Nevertheless, it is apparent that he did try to obtain a wide and diverse sample of teenagers, and it is not grossly inaccurate to consider this a cross-section of males within the specified age ranges, with perhaps a mild underrepresentation of the

TABLE 29

JOB SATISFACTION OF NORM GROUP

Item	% choosing item
a. It is exactly what I have wanted to do.	18
b. It is approximately what I have wanted to do.	28
c. It is something I entered, due to circumstances more or less beyond my control, but I am now satisfied in it.	40
d. It is a career which is tolerable but not really what I would like to do.	2
e. It is an unsatisfactory and unrewarding career.	1
f. It is a career that I strongly dislike and I wish I could leave for some other.	0.3

Note.—The instructions were as follows: Please check one of the following which best describes how you feel about your job.

TABLE 30

TEENAGE AND ADULT MEAN SCORES FOR BASIC SCALE NORM GROUP AND CORRELATIONS BETWEEN SET 1 AND SET 2 SCALES

Scale	Adult scores (age 52)[a]				Teenage scores (age 16)[b]		r
	Set 1		Set 2		Set 1		
	M	SD	M	SD	M	SD	
Public Speaking	−0.93	2.67	−1.41	3.94	46.6	9.2	.94
Law/Politics	−0.82	3.39	−1.41	4.08	45.4	9.6	.97
Business Management	0.43	5.03	1.03	6.01	43.9	9.5	.98
Sales	−2.58	4.15	−3.73	5.11	46.1	8.7	.96
Merchandising	−0.87	4.89	−0.58	6.54	44.1	9.6	.98
Office Practices	−1.63	3.63	−2.72	4.27	47.8	11.3	.98
Military Activities	−1.12	2.62	−1.02	3.04	50.9	9.4	.98
Technical Responsibility	0.56	2.08	0.56	2.08	44.5	11.3	1.00
Mathematics	2.39	3.35	2.39	3.35	45.4	10.2	1.00
Science	1.90	5.23	2.78	7.31	45.6	10.6	.98
Mechanical	4.54	10.97	5.40	12.07	47.5	9.1	1.00
Nature	2.67	5.06	3.08	5.44	43.3	10.3	.99
Agriculture	1.12	3.23	1.12	3.23	49.0	9.2	1.00
Adventure	−0.80	2.48	−0.23	3.53	57.2	11.1	.96
Recreational Leadership	0.80	3.66	1.92	4.86	52.9	9.7	.98
Medical Service	−0.02	3.28	−0.02	3.28	45.5	10.4	1.00
Social Service	−0.87	2.71	−0.90	3.35	47.3	8.7	.98
Religious Activities	−0.76	1.41	−1.18	3.79	46.9	9.2	.89
Teaching	−0.29	2.91	−0.17	3.65	40.8	9.0	.96
Music	−0.28	2.37	−0.63	4.68	46.6	10.1	.91
Art	−0.54	4.16	−1.28	6.63	46.0	9.3	.95
Writing	−0.66	5.09	−0.72	6.02	45.7	9.2	.98

[a] Raw scores. (The adult norm group has a standard score *M* of 50, *SD* of 10 on all scales.)
[b] Standard scores.

TABLE 31—BASIC SCALE INTERCORRELATIONS

Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1 Public Speaking		.72	.46	.25	.36	.10	.12	.04	−.16	−.22	−.18	.05	.09	.04	.28	.23	.35	.45	.37	.32	.21	.46
2 Law/Politics	.78		.62	.33	.53	.24	.12	.15	−.05	−.13	−.13	.01	.08	.19	.31	.25	.22	.20	.22	.15	.11	.37
3 Business Management	.58	.54		.67	.86	.63	.31	.41	−.06	−.17	.04	.00	.17	.24	.31	.17	.26	.22	.13	.02	−.04	.09
4 Sales	.51	.42	.69		.78	.63	.36	.29	−.10	−.15	.08	.11	.27	.22	.35	.13	.32	.32	.12	−.04	−.08	.01
5 Merchandising	.57	.49	.86	.80		.79	.27	.41	−.09	−.13	.11	.13	.21	.20	.31	.24	.28	.27	.14	.04	.03	.14
6 Office Practices	.30	.29	.56	.48	.66		.24	.38	.06	−.01	.24	.16	.16	.22	.18	.16	.29	.31	.20	.06	−.07	.06
7 Military Activities	.25	.21	.26	.17	.22	.19		.37	−.01	.14	.19	.21	.28	.36	.29	.21	.19	.20	.10	−.02	−.02	−.02
8 Technical Responsibility	.03	.10	.32	.16	.25	.23	.11		.24	.43	.65	.29	.31	.22	.08	.22	.06	.11	−.02	−.02	−.02	−.15
9 Mathematics	−.05	.06	.10	−.02	−.02	.08	.08	.26		.56	.43	−.05	−.13	.00	−.08	−.02	−.20	.11	−.11	−.11	−.16	−.27
10 Science	−.09	−.03	−.09	−.21	−.17	−.04	.04	.42	.53		.58	−.39	−.20	.07	−.08	.29	.01	.03	−.03	−.02	−.12	−.12
11 Mechanical	−.19	−.16	−.01	−.06	−.03	−.09	.12	.60	.37	.64		.33	.31	.17	−.06	.14	.02	.03	.06	.01	.13	−.20
12 Nature	−.04	−.12	−.13	−.15	−.06	−.02	.06	.18	.03	.40	.34		.76	.13	.19	.41	.26	.21	.02	.16	.34	.14
13 Agriculture	−.09	−.14	−.10	−.02	−.03	−.10	.09	.18	.01	.14	.27	.73		.29	.44	.28	.28	.20	.07	−.03	.14	−.01
14 Adventure	.21	.22	.11	.14	.09	−.19	.21	.25	.09	.20	.23	.08	.20		.45	.23	.08	.04	−.08	.02	.14	.04
15 Recreational Leadership	.23	.20	.23	.25	.28	.19	.24	.14	.04	−.05	.11	.07	.20	.34		.27	.36	.32	.15	.01	.10	.16
16 Medical Service	.19	.19	.13	.10	.13	.10	.12	.22	.19	.49	.31	.43	.22	.19	.15		.30	.21	.25	.19	.25	.13
17 Social Service	.49	.50	.35	.35	.39	.34	.16	.05	.02	.03	−.04	.12	.03	.03	.23	.25		.63	.45	.23	.24	.37
18 Religious Activities	.37	.19	.23	.22	.28	.31	.18	.15	.07	.13	.15	.27	.17	.00	.23	.17	.52		.43	.25	.08	.27
19 Teaching	.51	.45	.28	.16	.25	.25	.14	.08	.08	.28	.06	.26	.04	.08	−.13	.36	.32	.39		.62	.34	.51
20 Music	.31	.21	.12	.08	.15	.15	.05	.05	.03	.30	.11	.27	.03	.05	−.05	.34	.41	.42	.53		.50	.59
21 Art	.24	.18	.01	.04	.07	.00	.02	.07	−.04	.33	.16	.41	.08	.13	−.08	.39	.36	.24	.44	.68		.57
22 Writing	.54	.47	.26	.14	.26	.18	.10	.05	.02	.24	−.05	.21	−.03	.17	.07	.28	.45	.26	.63	.55	.59	

Note.—Set 1. Correlations above diagonal calculated on sample of Stanford University seniors, 1927 (N = 301).
Set 2. Correlations below diagonal calculated on cross-section sample of 52-yr-old men (N = 647).

lower end of the educational and socioeconomic ladder.

Some basic demographic data are available for the total group. Their educational level is listed in Table 28. Practically all (86%) are high school graduates; about two-thirds have had some college; over a third have degrees. One in eight have earned graduate degrees. Ninety-three percent were married in 1966. They had, on the average, 18-yr. experience in their jobs. Their feeling toward their work is reported in Table 29 which lists the percentage checking various choices in answer to a question on job satisfaction.

Raw-Score-to-Standard-Score Conversion

The mean scores for this norm group on both sets of scales are listed in Table 30. For the adults, only the raw-score statistics are listed; these are the figures that are used to convert an individual's raw scores into standard scores by using the following formula:

$$\frac{X - M}{SD}10 + 50 = \text{individual's standard score}$$

where X = individual's raw score, M = adult norm group raw score mean, SD = adult norm group raw score standard deviation. This conversion, which is a simple linear transformation and does not change the shape of the distribution, has been used on all the remaining scores reported in this report, including the teenager scores of the norm group. The net result is to convert all scores into a distribution where this adult norm group has a mean of 50 and standard deviation of 10. This permits an immediate comparison between the individual's score and the mean score of the norm group, and makes other comparisons possible also—such as comparing a person's score on the Law/Politics scale with his score on the Science scale.

The adult group was used here because both Set 1 and Set 2 scores were available for them; because the teenagers were tested in 1930 with the old booklet, only Set 1 scores are available for them, and the means are reported in standard score points. As can be seen, their means on almost all of the scales were slightly below the adult mean of 50. The only exceptions were the Adventure and Recreational Leadership scales.

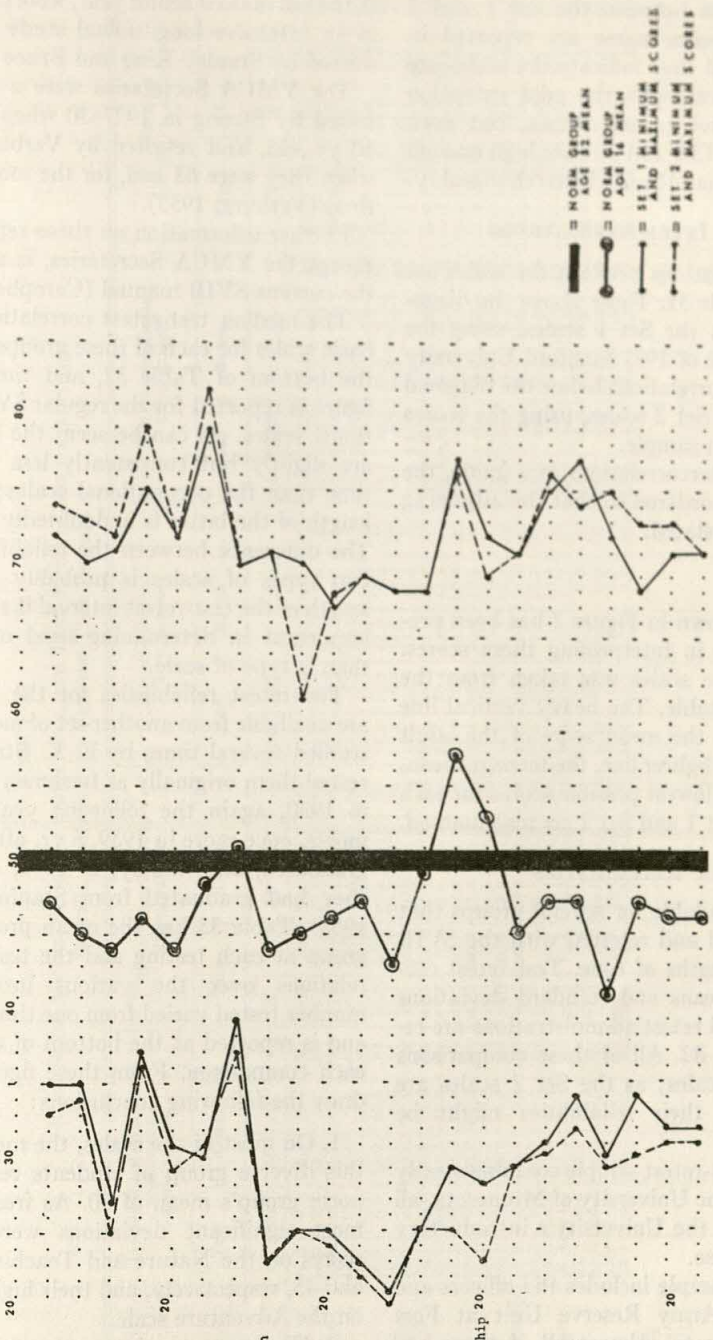


FIG. 2. A profile of basic scales for the SVIB—men.

The correlations between the Set 1 and 2 scales with the same name are reported in Table 30 also, and they indicate the scales are essentially identical, with the mild exception of Religious Activities and Music, but even their correlations (.89 and .91) are high enough so that the scales could be used interchangeably.

SCALE INTERCORRELATIONS

The intercorrelations between the scales are presented in Table 31. Those above the diagonal are based on the Set 1 scales, using the scores of a sample of 1927 Stanford University graduates. The correlations below the diagonal are based on the Set 2 scales, using the scores of the adult norm sample.

Using these intercorrelations as a guide, the scales have been ordered so that the adjoining ones tend to be related.

Suggested Profile

The profile shown in Figure 2 has been prepared as an aid in interpreting these scores. The order of the scales was taken from the intercorrelation table. The heavy vertical line at 50 represents the mean score of the adult norm group; the lighter line, the teenage mean. The highest and lowest possible scores for each scale for both Set 1 and Set 2 are also marked.

SCALE RELIABILITIES

Scores are available for several groups that have been tested and retested with the SVIB over varying lengths of time. Test-retest correlations and means and standard deviations for both test and retest administrations are reported in Table 32. All of these comparisons use the Set 1 scales; as the Set 2 scales are usually longer, their reliabilities might be slightly higher.

The 2-wk. test-retest sample contains mostly sophomores at the University of Minnesota, all volunteers from the University's introductory psychology course.

The 30-day sample includes the officers and men from an Army Reserve Unit at Fort Snelling, Minnesota. Almost all of them had college degrees. Their occupations were widely scattered, with a mild concentration in the advertising and mass media fields.

The Harvard students, who were tested in the spring of their freshmen year and retested in the fall of their senior year, were participants in an extensive longitudinal study being conducted by Stanley King and Bruce Finnie.

The YMCA Secretaries were a group first tested by Strong in 1927–30 when they were 40 yr. old, and retested by Verburg in 1951 when they were 63 and, for the most part, retired (Verburg, 1952).

Further information on these retest groups, except the YMCA Secretaries, is available in the current SVIB manual (Campbell, 1966).

The median test-retest correlation of these basic scales for each of these groups is given at the bottom of Table 32, and the analogous figure is reported for the regular SVIB occupational scales. As can be seen, the basic scales are slightly but consistently less stable over time than the occupational scales; the longer length of the latter is undoubtedly the reason. The difference between the reliabilities of the two types of scales is probably trivial; the length of the test-retest interval is still far more important in determining level of reliability than is type of scale.

Test-retest reliabilities for the basic scales are available from another set of men who were studied several times by E. K. Strong, Jr. He tested them originally as freshmen at Stanford in 1930, again the following year as sophomores, once more in 1939, 5 yr. after they had graduated, and, finally, in 1949, 15 yr. after they had graduated from Stanford (Strong, 1955). Table 33 has the mean profiles for the group at each testing and the test-retest correlations over the various intervals. The number tested varied from one time to another and is reported at the bottom of the table for each comparison. From these figures one can draw the following conclusions:

1. On most of the scales, the mean scores of this diverse group of students resembled the norm group's mean of 50. As freshmen, their most significant deviations were their low scores on the Nature and Teaching scales, 43 and 45, respectively, and their high score of 56 on the Adventure scale.

2. Nineteen years later, when retested, their interests had shifted mainly in these same three areas and, on the Nature and Adventure scales, they closely resembled the norm group. On the Teaching scale, however, the shift had carried them beyond the norm group to a score of 53.

TABLE 32

BASIC SCALE MEANS, STANDARD DEVIATIONS, AND TEST–RETEST CORRELATIONS FOR SEVERAL SAMPLES OVER VARYING TIME PERIODS

Scale	University of Minnesota sophomore 2-wk. sample					Army Reserve group 30-day sample					Harvard students 3-yr. sample					VMCA secretaries 22-yr. sample					Bankers 30-yr. sample				
	Test		Retest		r	Test		Retest		r	Test		Retest		r	Test		Retest		r	Test		Retest		r
	M	SD	M	SD		M	SD	M	SD		M	SD	M	SD		M	SD	M	SD		M	SD	M	SD	
Public Speaking	54.6	9.9	55.3	10.0	.89	58.0	9.4	58.7	9.5	.89	57.1	10.3	57.9	10.2	.71	59.6	7.8	62.3	6.5	.40	49.0	7.8	49.5	8.8	.30
Law/Politics	54.6	9.3	55.7	9.7	.88	57.1	10.2	57.2	10.1	.90	58.1	10.1	60.6	9.4	.50	52.6	8.1	55.7	9.3	.67	51.0	7.2	49.8	9.7	.39
Business Management	50.8	10.7	51.5	10.9	.90	52.8	9.5	54.1	9.3	.88	45.6	11.3	47.2	11.7	.72	58.1	7.4	58.2	8.6	.32	52.5	6.6	49.3	8.8	.35
Sales	49.9	9.5	50.6	9.8	.89	50.8	10.5	52.4	11.0	.89	45.3	8.9	45.2	8.8	.57	55.1	9.5	56.3	9.2	.63	52.8	9.0	51.5	9.6	.62
Merchandising	52.1	10.1	52.7	10.1	.89	52.7	10.6	53.8	10.6	.88	44.7	10.2	45.2	11.3	.66	55.3	7.5	55.4	8.2	.40	51.4	7.8	48.6	9.4	.24
Office Practices	50.3	9.7	52.2	10.5	.83	47.5	10.6	49.2	10.7	.87	42.9	10.1	42.6	10.4	.58	54.1	10.4	56.3	9.5	.50	60.2	8.3	58.9	9.4	.42
Military Activities	49.1	9.7	49.0	9.6	.91	50.5	9.3	50.7	9.6	.86	45.3	8.0	45.8	7.7	.52	50.6	9.7	50.1	8.7	.66	51.2	10.1	48.3	9.1	.59
Technical Supervision	49.2	9.6	49.6	10.1	.74	45.1	10.0	46.0	9.7	.71	43.2	10.9	45.0	12.4	.52	47.1	11.8	49.1	11.7	.61	45.0	10.1	42.4	10.5	.26
Mathematics	50.1	10.0	50.6	11.1	.94	47.5	11.3	47.9	10.8	.88	53.3	9.2	52.4	10.2	.71	49.1	10.5	49.3	10.1	.82	49.9	9.3	49.9	9.7	.58
Science	51.0	10.0	51.5	10.4	.94	47.1	9.6	47.4	10.3	.91	53.4	10.4	52.0	11.2	.70	49.3	9.1	47.1	9.2	.78	44.2	8.3	42.3	10.3	.67
Mechanical	47.2	10.0	48.3	10.7	.94	42.5	10.7	42.3	11.3	.94	42.4	9.7	45.1	10.9	.67	47.0	11.3	46.8	10.4	.79	45.1	11.2	42.9	11.0	.72
Nature	45.7	9.0	46.9	9.3	.90	42.2	10.3	42.5	10.1	.90	42.2	10.0	44.4	10.1	.68	51.1	9.7	52.3	8.5	.75	51.5	7.9	49.7	10.6	.61
Agriculture	46.3	9.9	47.1	10.2	.88	40.7	9.8	41.3	10.6	.89	42.0	10.2	43.7	10.2	.71	47.7	9.6	49.0	9.2	.73	49.2	7.8	49.0	8.6	.39
Adventure	58.1	8.5	59.4	9.1	.80	53.9	10.6	54.4	10.5	.81	56.0	11.5	57.2	12.2	.64	46.2	10.2	45.1	9.7	.53	43.8	9.3	39.3	7.3	.42
Recreational Leadership	51.8	9.6	53.1	9.6	.90	49.7	10.4	50.8	9.8	.92	47.2	10.8	48.9	11.3	.81	51.7	9.1	51.1	7.9	.72	46.4	8.3	42.8	9.1	.55
Medical Service	55.2	9.8	56.7	9.8	.89	50.7	9.5	51.1	9.3	.76	51.6	10.5	52.4	11.1	.67	53.6	8.7	52.1	8.2	.65	45.7	8.6	44.6	9.5	.38
Social Service	57.0	9.8	57.6	11.3	.83	55.6	9.7	54.7	9.5	.83	51.1	10.4	53.4	10.8	.54	70.3	5.4	71.7	3.9	.14	50.8	8.3	52.7	8.9	.51
Religious Activities	56.3	11.3	57.4	11.1	.83	53.5	10.8	54.3	10.1	.72	50.6	11.3	49.5	11.2	.51	70.6	6.8	72.2	6.5	.32	49.8	9.6	52.4	10.7	.51
Teaching	55.0	9.4	55.7	9.1	.88	54.2	8.4	55.1	8.3	.77	53.9	9.7	56.6	9.8	.63	55.6	8.6	58.1	8.0	.62	42.4	8.4	43.1	8.8	.41
Music	52.7	10.3	54.0	10.0	.89	53.5	8.5	54.0	8.7	.76	55.6	10.0	58.1	9.0	.60	53.8	7.9	53.5	8.6	.68	49.0	10.6	46.9	9.6	.66
Art	52.2	10.0	53.3	10.1	.93	54.8	9.3	54.4	10.2	.89	52.9	8.6	55.6	8.7	.56	53.9	8.1	51.8	7.9	.12	47.1	8.9	45.3	9.6	.56
Writing	53.4	9.3	54.7	9.5	.92	57.5	8.3	57.6	8.9	.85	59.7	8.0	61.7	7.5	.65	57.0	7.5	57.6	7.7	.72	46.5	8.1	46.1	8.2	.60
Mdn test-retest *r* for 22 basic scales					.89					.88					.65					.66					.53
Mdn test-retest *r* for 53 regular SVIB scales					.91					.91					.68					.61					.56

Note.—2-wk. sample $N = 140$; 30-day sample $N = 102$; 3-yr. sample $N = 189$; 22-yr. sample $N = 47$; 30-yr. sample $N = 48$.

TABLE 33

BASIS SCALE MEANS, STANDARD DEVIATIONS, AND TEST-RETEST CORRELATIONS FOR STANFORD STUDENTS TESTED AS FRESHMEN, SOPHOMORES, 5 YEARS AFTER GRADUATION, AND 15 YEARS AFTER GRADUATION

Scale	1930		1930–31	1931		1930–39	1931–39	1939		1930–49	1931–49	1939–49	1949	
	M	*SD*	*r*[a]	*M*	*SD*	*r*[b]	*r*[c]	*M*	*SD*	*r*[d]	*r*[e]	*r*[f]	*M*	*SD*
Public Speaking	52.2	9.8	.79	52.8	9.8	.51	.57	52.3	9.5	.54	.59	.66	52.1	9.7
Law/Politics	52.2	9.7	.79	53.8	10.1	.54	.56	53.6	9.1	.51	.56	.66	53.6	9.8
Business Management	49.2	10.3	.76	51.1	10.5	.54	.51	52.4	10.5	.48	.55	.78	51.9	10.2
Sales	47.9	8.7	.70	48.9	9.3	.48	.49	47.8	9.3	.46	.50	.81	48.5	9.6
Merchandising	50.3	10.2	.72	51.5	10.7	.52	.57	51.7	10.8	.51	.56	.81	51.3	10.6
Office Practices	51.2	10.7	.69	52.3	11.2	.48	.57	51.3	10.2	.46	.54	.72	49.7	10.5
Military Activities	51.9	9.9	.78	50.4	9.8	.63	.67	49.5	10.0	.56	.62	.74	49.0	9.6
Technical Supervision	47.8	11.8	.68	49.1	10.6	.51	.51	51.4	11.1	.51	.50	.69	51.4	11.8
Mathematics	51.1	9.9	.84	51.2	9.9	.75	.71	52.7	9.2	.73	.70	.87	52.1	9.5
Science	49.2	10.4	.82	49.1	10.0	.64	.72	52.4	10.0	.65	.72	.76	51.6	9.4
Mechanical	47.1	10.5	.85	46.7	10.3	.72	.77	48.6	11.2	.72	.70	.85	49.1	11.4
Nature	43.3	10.4	.81	43.4	10.6	.62	.68	47.2	10.2	.52	.53	.72	50.1	10.3
Agriculture	47.0	9.8	.77	46.9	9.8	.61	.67	47.5	9.8	.60	.60	.76	49.1	10.5
Adventure	56.1	11.2	.75	57.3	11.4	.54	.55	52.0	11.1	.59	.55	.68	50.1	10.8
Recreational Leadership	51.3	8.8	.77	51.9	9.1	.57	.68	49.9	9.5	.63	.67	.79	49.1	10.2
Medical Service	48.7	10.4	.80	48.3	9.9	.66	.66	49.8	11.1	.56	.61	.76	50.6	10.0
Social Service	49.2	9.7	.70	48.6	9.7	.34	.44	48.2	9.7	.38	.49	.77	49.0	9.9
Religious Activities	45.5	9.3	.68	44.3	9.5	.36	.37	44.7	9.7	.31	.31	.56	46.1	10.2
Teaching	44.8	9.6	.69	46.6	9.9	.45	.48	50.7	8.9	.40	.53	.62	52.7	9.2
Music	48.5	8.9	.75	49.0	9.2	.58	.51	51.3	8.1	.53	.51	.63	51.8	9.0
Art	47.9	9.3	.80	48.7	9.9	.59	.60	51.5	9.8	.56	.61	.79	51.2	10.1
Writing	51.4	9.8	.79	53.5	10.0	.67	.73	53.9	10.2	.64	.72	.81	54.9	9.6
Mdn test-retest *r*			.77			.56	.57			.54	.56	.76		

[a] $N = 253$.
[b] $N = 175$.
[c] $N = 159$.
[d] $N = 202$.
[e] $N = 178$.
[f] $N = 137$.

3. If one assumes that a 3-point shift—which is roughly one-third *SD*—is the smallest difference worth attending to, the following shifts occurred over time within this sample:

GAINS

Teaching	8 points
Nature	7 points
Music	4 points
Writing	4 points
Art	3 points
Business Management	3 points
Technical Supervision	3 points

LOSSES

Adventure	6 points
Military Activities	3 points

"Middle-aging intellectual responsibility" is the apparent trend.

4. Their scores on the remaining scales over this 19-yr. period did not shift even this mild amount. Interests in areas such as Public Speaking, Mathematics, Science, and Religious Activities were, as measured by these scales, surprisingly constant.

5. The test-retest correlations reported at the bottom of Table 33 provide a measure of the stability of the rank order of the individuals within the total sample. The most intriguing finding here is that the stability over the 10-yr. adult span, 1939–1949, was equal to the 1-yr. freshmen-sophomore period—further support for the familiar finding that interests are remarkably stable after age 25.

Similar results for another group of college students are reported in the following section on validity.

VALIDITY

There is increasing awareness among psychologists that there is no single way of establishing the validity of any test or inventory as no one index can adequately represent the relationship between test behavior and real-world behavior. Even in those situations where the test is designed for a specific purpose, for example, the Medical School Aptitude Test, it is not clear which aspect of medical school performance should be used as the criterion. For measures of interests, the situation is even more confusing as there are no criteria available that have even the relative clarity of medical school accomplishments. How can the validity of, say,

TABLE 34

BRIEF DESCRIPTION OF THE OCCUPATIONAL SAMPLES WHOSE BASIC SCALE MEAN SCORES APPEAR IN TABLES 35–56

Sample	*N*	Year tested	Investigator	Brief description
Accountant	345	1932	Strong	Junior accountants
Advertising Men	169	1931	Strong	
Air Force Officer	198	1960	Winter	Air Force officers training at Wright Patterson Air Force Base
Animal Husbandry Prof.	82	1967	CIMR[a]	University faculty members
Anthropologist	161	1967	Rossmann & Lips	University faculty
Army officer	463	1950	Army Personnel Research Office	Above average officers, mostly field grade, all West Point graduates
Architect	240	1933	Strong	Architects, mostly from California
Artist	231	1933	Strong	
Astronaut	16	1966	CIMR	Astronauts, mostly military officers
Authors	298	1931	Strong	Roughly 60% journalists, 40% authors
Banker ('34)	238	1934	Strong	
Banker ('64)	98	1964	CIMR	
Biologist	342	1959	Lindsay	All listed in *American Men of Science*
Bus. Ed. Teacher	323	1956	Bacon	
Carpenter	181	1936	Strong	
CPA ('44)	611	1944	Strong	Senior Certified Public Accountants
CPA ('65)	304	1965	Rhode	Minnesota Certified Public Accountants
Cham Comm Exec	400	1960	Strong	Chamber of Commerce and Trade Association executives
Chemist	297	1931	Strong	Industrial chemists
Comm Rec Director	350	1964	Roys	Community recreation directors
Computer Programmer	500	1964	Perry & Cannon	
Corp Pres ('35)	169	1935	Strong	Corporation presidents
Corp Pres ('65)	25	1965	CIMR	Corporation presidents
Credit Manager	452	1958	Strong	
Danforth Fellow	93	1966	Rossmann & Bentley	Danforth Foundation Fellows—aspiring college professors
Dentist	239	1932	Strong	
Economist	99	1967	Rossmann & Lips	University faculty members
Elem Teacher	115	1965	Winkle	Minnesota elementary teachers
Engineer	513	1928	Strong	
Farmer ('36)	241	1936	Strong	Farmers, mostly from Pacific Northwest
Farmer ('67)	77	1967	Krause	South Dakota farmers
Forest Service Men	406	1936	Strong	
Funeral Director	360	1945	Strong	
Generals and Admirals	51	1965	CIMR	3- and 4-Star Army Generals and Navy Admirals
Governor	28	1965	CIMR	Retired state governors
Guidance Counselor	275	c1950	Brown	
Interior Decorator	190	1967	CIMR	Members, American Institute of Interior Design
Judge	42	1967	CIMR	Federal District Judges
Lawyer ('27)	251	1927	Strong	California lawyers
Lawyer ('49)	78	1949	Berdie & Hagenah	Minnesota lawyers
Legislator	81	1967	Willow	Senators and Representatives from Minnesota Legislature
Librarian	425	1959	Winters	
Mathematician	181	1929	Strong	
Math-Science Teacher	289	1936	Strong	High school mathematics-science teachers
McKnight Fellow	32	1966	CIMR	McKnight Fellows—actors in Tyrone Guthrie Theater in Minneapolis
Minister ('27)	250	1927	Strong	

[a] CIMR = Center for Interest Measurement Research, University of Minnesota.

TABLE 34—(Continued)

Sample	*N*	Year tested	Investigator	Brief description
Minister ('65)	96	1965	CIMR	
Mpls. Symphony	53	1966	CIMR	Members, Minneapolis Symphony Orchestra
Musician	450	1952	Strong	Mainly performing musicians
Music Teacher	493	1952	Strong	
NIAL Member	42	1966	CIMR	Members, National Institute of Arts and Letters
Navy Officer	645	1965	Stephenson & Abrahams	Above average naval officers
Newsmen	283	1967	CIMR	Newspaper reporters and television news broadcasters
Office Worker	326	1928	Strong	
Optometrist	405	1963	Acree	Members, American Optometric Association
Osteopath	585	1939	Strong	
Personnel Director	147	1927	Strong	
Petroleum Engineer	385	1965	Alford	Texas petroleum engineers
Pharmacist	309	1947	Schwebel	New York pharmacists
Photographer	253	1967	CIMR	Mostly news photographers and photojournalists
Physical Therapist	350	1957	Strong	
Physician ('27)	175	1927	Strong	
Physician ('49)	534	1949	Strong & Tucker	Random sample of the American Medical Association
Internist	209	1949	Strong & Tucker	Representative sample of those boarded in designated specialty
Neurological Surgeon	47	1949	Strong & Tucker	
Orthopedic Surgeon	71	1949	Strong & Tucker	Representative sample of those boarded in designated specialty
Pathologist	154	1949	Strong & Tucker	Representative sample of those boarded in designated specialty
Pediatrician	96	1949	Strong & Tucker	Representative sample of those boarded in designated specialty
Physiatrist	460	1966	Athelstan	Representative sample of those boarded in designated specialty
Psychiatrist	404	1949	Strong & Tucker	Representative sample of those boarded in designated specialty
Radiologist	111	1949	Strong & Tucker	Representative sample of those boarded in designated specialty
Surgeon	188	1949	Strong & Tucker	Representative sample of those boarded in designated specialty
Urologist	84	1949	Strong & Tucker	Representative sample of those boarded in designated specialty
Physicist ('27)	192	1929	Strong	
Physicist ('67)	64	1967	CIMR	
Pilot	510	1941	Strong	Strong's original aviator group
Policemen	254	1933	Strong	Police officers from several cities
Political Scientist	177	1967	Rossmann & Lips	University faculty members
Priest	234	1966	Lepak	Minnesota priests
Printer	270	1936	Strong	
Production Manager	216	1935	Strong	
Psychologist ('31)	174	1931	Strong	
Psychologist ('47)	149	1947	Kelly	Veterans Administration clinical psychologists
Psychologist ('49)	1045	1949	Kriedt	Ninety percent of APA membership
Psychologist (Exp)	256	1949	Kriedt	Experimental psychologists, subset of the total 1949 group
Public Admin	550	1941	Strong	Public administrators from a variety of public agencies

TABLE 34—(Continued)

Sample	N	Year tested	Investigator	Brief description
Pulitzer Prize	84	1965	CIMR	Pulitzer Prize winners, mostly journalists and novelists
Purch Agent	219	1931	Strong	Purchasing agents
Rehab Counselor	272	1949	Acree	Vocational rehabilitation counselors from the Veterans Administration
Salesmen, Auto	116	1929	Strong	Automobile salesmen
Salesmen, Computer	190	1964	Stein	Control Data Corporation computer salesmen
Salesmen, Encyclopedia	49	1966	CIMR	Outstanding World Book encyclopedia salesmen
Salesmen, Life ('31)	310	1931	Strong	
Salesmen, Life ('66)	76	1966	CIMR	$2,000,000 life insurance salesmen
Salesmen, 3M	100	c1960	Kirchner	Salesmen on staff of 3M
Salesmen, 3M applicants	100	c1960	Kirchner	Men applying for sales positions at 3M company
Salesmen, PG&E	179	1939	Strong	Appliance salesmen, Pacific Gas & Electric company
Salesmen, Real Estate	243	1932	Strong	
Salesmen, Steel	61	1966	Berdie	Structural steel salesmen
Sales Manager	228	1932	Strong	
School Super ('30)	190	1930	Strong	School superintendents
School Super ('65)	153	1965	CIMR	School superintendents
Soc Sci Teacher	217	1936	Strong	High school social science teachers
Social Worker ('53)	400	1953	McCornack	
Social Worker ('67)	54	1967	CIMR	
Sociologist	198	1965	Rossmann & Lips	University faculty members
Student Personnel	192	1961	Clark	Members, American College Personnel Association
Unitarian Minister	113	1950	Strong	
Vet ('49)	310	1949	Hannum	Iowa veterinarians
Vet ('66)	478	1966	Hannum & Alsip	Iowa veterinarians
YMCA PD	216	1927	Strong	YMCA phyiscal director
YMCA Sec	113	1927	Strong	
YMCA Staff	184	1961	Seashore	YMCA staff (comparable to combination of original physical directors and secretaries)

a measure of interests in Religious Activities be studied?

The establishment of validity for such measures, it is the authors' opinion, depends on the collection and organization of a vast amount of data covering several aspects of the inventory. All psychological tests and inventories have some potential validity—for persons do not give random answers—but until the relationships between the test behavior and an individual's actions are thoroughly studied, using many diverse approaches, the value of the test remains obscure.

With this orientation, a substantial amount of information is presented here, showing that responses to these SVIB scales are related to the occupation that the individual chooses. The concept of occupational membership was used here because it has proven valuable in earlier work with the regular SVIB scales, because it is objective, because—as E. K. Strong frequently commented—hardly anyone finds fault with it, and because it is closely related to the intended purpose of the SVIB, that is, as an aid in vocational counseling.

The information that follows has been organized in three traditional areas: content, concurrent, and predictive validity.

Content Validity

Content validity refers to the actual content of the scales. For these basic scales, a certain amount of content validity is assured as each item has survived both a statistical and a "common sense" screening to determine that, within each scale, the items are related. Con-

TABLE 35

MALE OCCUPATIONS MEAN SCORES ON THE ADVENTURE SCALE

M standard score	Sample
70	
69	
68	
67	
66	Astronaut
65	
64	
63	
62	
61	
60	Salesmen, 3M applicants
59	Salesmen, 3M/ Policemen/ Salesmen, Computer
58	Air Force Officer/ Navy Officer
57	Salesmen, Steel/ Pilot
56	McKnight Fellow/ YMCA Staff/ Petroleum Engineer
55	Photographer/ Elem Teacher/ Comm Rec Director/ Computer Programmer
54	CPA 65/ Army Officer/ Legislator/ Newsmen/ Salesmen, Life 66/ Social Worker 67
53	Cham Comm Exec/ Salesmen, Auto/ Forest Service Men/ Physical Therapist
52	Vet 66/ YMCA PD/ Optometrist/ Minister 65/ Orthopedic Surgeon
51	Salesmen, Encyclopedia/ Generals and Admirals/ Danforth Fellow/ CPA 44/ Corp Pres 65/ Neurological Surgeon
50	Priest/ Sociologist/ Credit Manager/ Salesmen, PG&E/ Printer/ Carpenter/ School Super 65
49	Radiologist/ Urologist/ Farmer 36/ Surgeon/ Mpls Symphony/ Osteopath/ Math-Science Teacher/ Student Personnel
48	Psychologist 47/ Salesmen, Real Estate/ Public Admin/ Pediatrician/ Chemist/ Biologist/ Banker 64/ Physiatrist/ Bus Ed Teacher/ Farmer 67/ Musician/ Physicist 67/ Anthropologist
47	Accountant/ Social Worker 53/ Psychologist 31/ Office Worker/ Unitarian Minister/ Economist/ Production Manager/ Psychologist (Exp)/ Personnel Director/ Funeral Director/ Physician 49/ Psychiatrist/ Animal Husbandry Prof/ Artist/ Vet 49
46	Pharmacist/ Soc Sci Teacher/ Physician 27/ Engineer/ Purch Agent/ Political Scientist/ Sales Manager/ Librarian/ Dentist/ Lawyer 49/ Lawyer 27/ Pathologist/ Authors/ Interior Decorator/ Governor/ Advertising Men
45	Rehab Counselor/ Architect/ YMCA Sec/ Salesmen, Life 31/ Guidance Counselor/ Pulitzer Prize/ Minister 27/ Psychologist 49/ Music Teacher/ Internist
44	Judge
43	Banker 34/ Corp Pres 35/ Physicist 27
42	
41	School Super 30
40	Mathematician/ NIAL Member
39	
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

sequently one has some confidence that when a person answers, for example, Like or Dislike to the items, Algebra, Arithmetic, Calculus, Geometry, Mathematics, and Physics, he is telling something about his interests in mathematics.

However, although content validity is quite helpful in scale interpretation, by itself it reveals nothing about the relationship between an individual's interests and his actual behavior. For such information, one needs to turn to concurrent and predictive validity.

Concurrent Validity

Concurrent validity refers to the power of the test to distinguish between specified groups

TABLE 36

MALE OCCUPATIONS MEAN SCORES ON THE AGRICULTURE SCALE

M standard score	Sample
70	
69	
68	
67	
66	
65	
64	
63	Farmer 67
62	
61	Farmer 36/ Vet 49
60	Vet 66
59	Forest Service Men/ Animal Husbandry Prof
58	
57	
56	
55	
54	Pilot
53	Army Officer/ Generals and Admirals
52	Orthopedic Surgeon
51	Salesmen, Encyclopedia/ Policemen/ Comm Rec Director/ Physical Therapist/ Astronaut/ Governor
50	YMCA PD/ Salesmen, PG&E/ Biologist/ Urologist/ Legislator/ Osteopath
49	Public Admin/ Radiologist/ Carpenter/ Surgeon/ Salesmen, Life 66/ Petroleum Engineer/ Neurological Surgeon
48	Salesmen, Real Estate/ YMCA Sec/ Salesmen, Steel/ Salesmen, Life 31/ YMCA Staff/ Dentist/ Banker 34/ Minister 27/ Salesmen, Computer/ Funeral Director/ Air Force Officer/ Navy Officer/ Physiatrist
47	Salesmen, 3M applicants/ Salesmen, 3M/ Soc Sci Teacher/ Pediatrician/ Physician 27/ Credit Manager/ Photographer/ Elem Teacher/ Sales Manager/ CPA 44/ Banker 64/ Corp Pres 65/ Physician 49/ Artist
46	Chemist/ Engineer/ School Super 30/ Unitarian Minister/ Production Manager/ Personnel Director/ Salesmen, Auto/ Corp Pres 35/ Bus Ed Teacher/ Pathologist/ Math-Science Teacher
45	Rehab Counselor/ Priest/ Cham Comm Exec/ Architect/ Social Worker 53/ Office Worker/ Printer/ Guidance Counselor/ McKnight Fellow/ Optometrist/ Computer Programmer/ Lawyer 49/ School Super 65/ Psychiatrist/ Social Worker 67/ Music Teacher/ Musician/ Internist/ Anthropologist
44	Pharmacist/ Accountant/ Psychologist (Exp)/ Purch Agent/ Lawyer 27/ Newsmen/ Minister 65/ Advertising Men/ Physicist 27/ Psychologist 49/ Authors
43	Psychologist 31/ Mpls Symphony/ Danforth Fellow/ Physicist 67/ Interior Decorator/ NIAL Member/ Student Personnel
42	Psychologist 47/ Sociologist/ CPA 65/ Librarian/ Pulitzer Prize/ Judge/ Mathematician
41	Economist
40	Political Scientist
39	
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

at the current moment. For example, can these scales discriminate between the interests of men who *are* currently salesmen or scientists? This is in contrast to predictive validity where the issue is whether the scales can identify differences between boys who *will become* scientists or salesmen.

The relevant data for concurrent validity are presented in Tables 34 through 56. For each of the basic scales, the means of 113 occupational samples are rank ordered, thus permitting at a glance an evaluation of the power of the scale to separate occupations.

These occupational samples represent the

TABLE 37

MALE OCCUPATIONS MEAN SCORES ON THE ART SCALE

M standard score	Sample
70	
69	
68	
67	
66	
65	Interior Decorator
64	Architect/ McKnight Fellow
63	Artist
62	
61	
60	Photographer
59	Mpls Symphony/ NIAL Member
58	Pulitzer Prize
57	Unitarian Minister/ Librarian/ Danforth Fellow/ Musician
56	Advertising Men/ Physiatrist/ Music Teacher/ Anthropologist/ Neurological Surgeon
55	Psychologist 47/ Psychologist 31/ Elem Teacher/ Minister 27/ Minister 65/ Orthopedic Surgeon/ Authors
54	Priest/ Sociologist/ Social Worker 53/ Pediatrician/ Biologist/ Economist/ Psychologist (Exp)/ Dentist/ Newsmen/ Psychiatrist
53	Salesmen, 3M/ YMCA Sec/ Urologist/ Personnel Director/ Political Scientist/ Judge/ Physical Therapist/ Psychologist/ 49/ Social Worker 67/ Physicist 67/ Pathologist
52	Salesmen, 3M applicants/ YMCA PD/ Radiologist/ Physician 27/ Salesmen, PG&E/ Carpenter/ Surgeon/ Salesmen, Encyclopedia/ Comm Rec Director/ YMCA Staff/ Computer Programmer/ Generals and Admirals/ Physicist 27/ Physician 49/ Internist/ Student Personnel
51	Rehab Counselor/ Cham Comm Exec/ Public Admin/ Office Worker/ Engineer/ Printer/ Corp Pres 35/ Optometrist/ Bus Ed Teacher/ Osteopath/ Astronaut/ Mathematician
50	Accountant/ Soc Sci Teacher/ Chemist/ Credit Manager/ Salesmen, Life 31/ School Super 30/ Guidance Counselor/ Funeral Direc/ Policemen/ Army Officer/ Legislator/ School Super 65/ Salesmen, Life 66/ Math-Science Teacher
49	Salesmen, Real Estate/ Salesmen, Steel/ Air Force Officer/ Production Managers/ Salesmen, Auto/ Forest Service Men/ Pilot/ Sales Manager/ Salesmen, Computer/ APA 44/ Navy Officer/ Lawyer 27/ Governor
48	Pharmacist/ Farmer 36/ Purch Agent/ Lawyer 49/ Animal Husbandry Prof/ Petroleum Engineer
47	Vet 66/ CPA 65/ Banker 34/ Banker 64/ Corp Pres 65
46	Vet 49
45	Farmer 67
44	
43	
42	
41	
40	
39	
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

majority of adult men who have been tested for research purposes with the SVIB since it was first published in 1927. Further information on each sample is listed in Table 57.

Most of these occupational samples include men with at least 3 yr. of experience who say they like their work. More extensive information on many of the samples can be found in the SVIB manual (Campbell, 1966).

One large block of these samples came from Strong's files, including those used to build the occupational scales, another block came from

TABLE 38

MALE OCCUPATIONS MEAN SCORES ON THE BUSINESS MANAGEMENT SCALE

M standard score	Sample
70	
69	
68	
67	
66	
65	
64	
63	
62	Salesmen, 3M applicant
61	Salesmen, 3M/ Credit Manager
60	Cham Comm Exec/ Sales, Encyclopedia/ Bus Ed Teacher
59	YMCA Sec/ Personnel Director/ Salesmen, Computer
58	Salesmen, Steel/ Comm Rec Director
57	Rehab Counselor/ Accountant/ Office Worker/ Salesmen, PB&E/ Sales Manager/ YMCA Staff/ School Super 65
56	Guidance Counselor/ Purch Agent/ Legislator
55	Public Admin/ Corp Pres 35/ Funeral Director/ CPA 65/ Production Manager/ Army Officer/ CPA 44
54	Soc Sci Teacher/ Social Worker 53/ Salesmen, Life 31/ Air Force Officer/ Salesmen, Auto/ Navy Officer/ Banker 64/ Governor
53	Salesmen, Real Estate/ School Super 30/ Banker 34/ Generals and Admirals/ Corp Pres 65/ Salesmen, Life 66
52	YMCA PD/ Petroleum Engineer/ Student Personnel
51	Pharmacist/ Elem Teacher/ Forest Service Men/ Social Worker 67/ Music Teacher
50	Policemen/ Optometrist/ Lawyer 49/ Minister 65/ Math-Science Teacher/ Advertising Men
49	Psychologist 47/ Priest/ Librarian/ Computer Programmer/ Minister 27/ Physical Therapist
48	Vet 66/ Engineer/ Printer/ Unitarian Minister/ Carpenter/ Farmer 36/ Pilot/ Interior Decorator
47	Economist/ Physiatrist/ Psychologist 49/ Orthopedic Surgeon/ Vet 49
46	Pediatrician/ Chemist/ Urologist/ Lawyer 27/ Newsmen/ Psychiatrist/ Farmer 67/ Osteopath/ Astronaut
45	Sociologist/ Radiologist/ Judge/ Animal Husbandry Prof/ Musician
44	Psychologist 31/ Surgeon/ Political Scientist/ Dentist/ Neurological Surgeon
43	Architect/ Psychologist (Exp)/ Physician 49/ Internist
42	Biologist/ Danforth Fellow
41	Physician 27/ Photographer/ Mpls Symphony/ Pathologist
40	McKnight Fellow/ Physicist 27/ Physicist 67/ Mathematician/ Authors
39	Anthropologist
38	Pulitzer Prize
37	
36	Artist/ NIAL Member
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

the work of the present authors' research institute at Minnesota, and the remainder came from various other investigators working with the SVIB. Practically everyone in the latter group who was asked for his data was cooperative—if the information had not been lost or discarded—and the authors would like to publicly acknowledge gratitude for their help.

The mean scores in Tables 35 through 56 are reassuring. Occupations which should score high did so; military officers scored highest on the Military Activities scale, ministers scored highest on the Religious Activities scale, scientists scored highest on the Science scale and so on—there were virtually no surprises. In fact, the results are so straightforward as to lead one to that conclusion voiced frequently

TABLE 39

MALE OCCUPATIONS MEAN SCORES ON THE LAW/POLITICS SCALE

M standard score	Sample
70	
69	
68	
67	
66	
65	
64	
63	
62	Judge/ Legislator
61	Salesmen, 3M applicants/ Salesmen, 3M/ Political Scientist/ Salesmen, Computer/ Governor
60	Cham Comm Exec
59	Laywer 49
58	Social Worker 53/ Credit Manager/ CPA 65/ Economist/ Navy Officer/ Lawyer 27/ Newsmen/ Salesmen, Life 66/ Student Personnel
57	Rehab Counselor/ Unitarian Minister/ Air Force Officer/ Salesmen, Encyclopedia/ Comm Rec Director/ Army Officer/ School Super 65/ Social Worker 67
56	Priest/ Sociologist/ Salesmen, Steel/ Guidance Counselor/ Personnel Director/ YMCA Staff/ Generals and Admirals/ Bus Ed Teacher/ Danforth Fellow
55	Public Admin/ Social Sci Teacher/ School Super 30/ Elem Teacher/ Minister 65/ Astronauts
54	Psychologist 47/ Policemen/ Optometrist/ Computer Programmer/ Corp Pres 65/ Petroleum Engineer
53	YMCA Sec/ Salesmen, PG&E/ McKnight Fellow/ CPA 44/ Physiatrist/ Physical Therapist/ Psychiatrist/ Psychologist 49
52	Accountant/ Pediatrician/ Office Worker/ Salesmen, Life 31/ Salesmen, Auto/ Librarian/ Banker 34/ Minister 27/ Orthopedic Surgeon/ Anthropologist/ Neurological Surgeon
51	Vet 66/ Salesmen, Real Estate/ Psychologist 31/ Radiologist/ Urologist/ Psychologist (Exp)/ Pulitzer Prize/ Banker 64
50	Printer/ Production Manager/ Purch Agent/ Pilot/ Sales Manager
49	Pharmacist/ YMCA PD/ Corp Pres 35/ Biologist/ Photographer/ Surgeon/ Forest Service Men/ Animal Husbandry Professor/ Music Teacher/ Osteopath/ Internist/ Math-Science Teacher
48	Funeral Director/ Mpls Symphony/ Advertising Men/ Physician 49/ Musician/ Physicist 67/ Authors
47	Engineer/ Farmer 36/ Dentist/ Pathologist/ Vet 49
46	Physician 27/ Chemist/ Carpenter/ Mathematician/ Interior Decorator/ NIAL Member
45	Architect/ Farmer 67
44	Physicist 27
43	
42	Artist
41	
40	
39	
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

by students in introductory psychology courses: "You had to do all that research just to find *that* out?"

Clearly, scores on these scales are related, in a highly significant fashion, to the occupations of adult men. Anyone who would understand these scales thoroughly must spend some time studying these tables of means, both to learn more about the relative levels of scores on each scale, and to learn more about each occupation.

These rank-ordered means provide a firmer foundation for ascertaining the importance of scores at various levels. Scanning these tables suggests that scores above 58 or 59 are high

TABLE 40

MALE OCCUPATIONS MEAN SCORES ON THE MATHEMATICS SCALE

M standard score	Sample
70	
69	
68	
67	
66	
65	
64	
63	
62	
61	Astronauts
60	Physicist 27/ Physicist 67/ Mathematician
59	
58	Air Force Officer/ Computer Programmer/ Salesmen, Computer/ Petroleum Engineer
57	Engineer/ Army Officer
56	Chemist/ CPA 65/ Economist/ Generals and Admirals/ Corp Pres 65
55	Psychologist (Exp)/ Optometrist/ CPA 44/ Navy Officer/ Animal Husbandry Prof
54	Psychologist 31/ Danforth Fellow/ Physiatrist/ Math-Science Teacher
53	Accountant/ Pediatrician/ School Super 30/ Biologist/ Urologist/ Production Manager/ Elem Teacher/ Pilot/ School Super 65/ Psychologist 49/ Pathologist/ Orthopedic Surgeon/ Neurological Surgeon
52	Architect/ Radiologist/ Credit Manager/ Salesmen, Steel/ Corp Pres 35/ Forest Service Men/ Banker 64/ Psychiatrist
51	Salesmen, 3M/ Public Admin/ Office Workers/ Personnel Director/ Surgeon/ Sales Manager/ Lawyer 49/ Physician 49/ Internist
50	Psychologist 47/ Vet 66/ Pharmacist/ Sociologist/ Salesmen, PG&E/ Guidance Counselor/ Carpenter/ Farmer 36/ Purch Agent/ Judge/ Legislator/ Physical Therapist/ Bus Ed Teacher/ Student Personnel
49	Rehab Counselor/ Physician 27/ Salesmen, Auto/ Banker 34/ Lawyer 27/ NIAL Member/ Governor
48	Salesmen, 3M applicant/ YMCA PD/ Policemen/ Political Scientist/ Mpls Symphony/ Comm Rec Director/ Dentist/ Minister 27/ Osteopath/ Anthropologist/ Vet 49
47	Cham Comm Exec/ YMCA Sec/ Salesmen, Life 31/ Printer/ Unitarian Minister/ Funeral Director/ Salesmen, Encyclopedia
46	Priest/ Salesmen, Real Estate/ Photographer/ McKnight Fellow/ Librarian/ YMCA Staff/ Minister 65/ Salesmen, Life 66/ Music Teacher/ Musician
45	Soc Sci Teacher/ Social Worker 53/ Social Worker 67/ Farmer 67/ Advertising Men
44	
43	Newsmen
42	Pulitzer Prize/ Artist/ Interior Decorator
41	
40	Authors
39	
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

enough to be important from an occupational standpoint; scores below 42 or 43 indicate important areas of rejection, at least among adults. Thus, the present authors' recommendation is that scores above 58 be considered high, those below 42, low. Some sharpening of this interpretation will obviously be necessary, especially with teenagers and especially with the Adventure scale where one knows high scores will likely decrease over time, but these figures can be used as initial landmarks.

Though these data are positive and meaning-

TABLE 41

MALE OCCUPATIONS MEAN SCORES ON THE MECHANICAL SCALE

M standard score	Sample
70	
69	
68	
67	
66	
65	
64	
63	
62	
61	
60	
59	
58	
57	
56	Air Force Officer/ Carpenter/ Orthopedic Surgeon
55	
54	Radiologist/ Pilot/ Army Officer/ Farmer 67/ Neurological Surgeon
53	Engineer/ Urologist/ Physicist 27/ Petroleum Engineer
52	Psychologist (Exp)/ Computer Programmer/ Physical Therapist/ Astronaut/ Math-Science Teacher
51	Chemist/ Printer/ Production Manager/ Surgeon/ Forest Service Men/ Policemen/ Navy Officer/ Physiatrist/ Physicist 67
50	Pediatrician/ Salesmen, PG&E/ Farmer 36/ Optometrist/ Salesmen, Computer/ Generals and Admirals/ Pathologist
49	Biologist/ Dentist/ CPA 44/ Psychiatrist/ Osteopath
48	Architect/ YMCA PD/ Psychologist 31/ Credit Manager/ Personnel Director/ Elem Teacher/ Physician 49/ Psychologist 49
47	Vet 66/ Public Admin/ Accountant/ Salesmen, Auto/ Purch Agent/ Comm Rec Director/ Minister 27/ Banker 64/ Corp Pres 65/ Animal Husbandry Prof/ Bus Ed Teacher/ Internist/ Mathematician/ Vet 49
46	Salesmen, 3M applicants/ Salesmen, 3M/ YMCA Sec/ Physician 27/ Salesmen, Steel/ Office Worker/ Unitarian Minister/ Guidance Counselor/ Corp Pres 35/ Mpls Symphony/ YMCA Staff/ Musician/
45	Psychologist 47/ Rehab Counselor/ Pharmacist/ Social Worker 53/ CPA 65/ Photographer/ Social Worker 67
44	Priest/ Cham Comm Exec/ Soc Sci Teacher/ School Super 30/ Funeral Director/ Sales Manager/ Legislator/ Banker 34/ Minister 65/ School Super 65/ Music Teacher/ Artist
43	Salesmen, Real Estate/ Sociologist/ Economist/ Librarian/ Anthropologist/ Student Personnel
42	Danforth Fellow
41	Salesmen, Encyclopedia/ McKnight Fellow/ Lawyer 49/ Advertising Men/ Governor
40	Salesmen, Life 31/ Lawyer 27/ Interior Decorator
39	Judge/ Newsmen
38	Political Scientist/ Pulitzer Prize/ Salesmen, Life 66/ Authors/ NIAL Member
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

ful, one disappointing aspect—at least for the senior author—was the magnitude of differences between extreme groups. On most of the scales, the range from the highest mean to the lowest was about 20 standard score points, or 2 *SD*s. While that is a substantial separation, especially when viewed against results from other tests (if reported in percentiles, that would be the difference between the fifteenth and eighty-fifth percentiles), still it is less than the 3 or 4 *SD*s found between extreme scores on the regular SVIB occupational scales. To achieve purity of content, and thus easier interpretation, empirical validity has suffered.

TABLE 42

MALE OCCUPATIONS MEAN SCORES ON THE MEDICAL SERVICE SCALE

M standard score	Sample
70	
69	
68	
67	
66	
65	
64	Dentist/ Physical Therapist
63	Vet 66/ Urologist/ Physiatrist/ Osteopath/ Vet 49
62	Pharmacist/ Radiologist/ Pediatrician/ Surgeon/ Physician 49/ Neurological Surgeon
61	Physician 27
60	Optometrist/ Psychiatrist/ Pathologist/ Internist/ Orthopedic Surgeon
59	YMCA PD/ Animal Husbandry Prof
58	Biologist
57	Psychologist 47/ Math-Science Teacher
56	Psychologist (Exp)/ Comm Rec Director/ YMCA Staff
55	Psychologist 31/ Psychologist 49
54	Salesmen, 3M applicants/ Rehab Counselor/ Priest/ Social Worker 53/ Chemist/ Elem Teacher/ Policemen/ Minister 27/ Social Worker 67/ Astronaut
53	YMCA Sec/ Unitarian Minister/ Guidance Counselor/ Funeral Director/ Salesmen, Computer/ School Super 65/ Student Personnel
52	Salesmen, 3M/ Soc Sci Teacher/ Credit Manager/ Air Force Officer/ Mpls Symphony/ Army Officer/ Computer Programmer/ Legislator/ Danforth Fellow/ Minister 65/ Salesmen, Life 66/ Petroleum Engineer
51	Cham Comm Exec/ Public Admin/ Salesmen, PG&E/ Forest Service Men/ Pilot/ Navy Officer/ Musician/ Governor
50	Sociologist/ Salesmen Steel/ Office Worker/ School Super 30/ Personnel Director/ Farmer 36/ Salesmen, Auto/ Librarian/ CPA 44/ Music Teacher/ Bus Ed Teacher/ Anthropologist
49	Salesmen, Life 31/ Printer/ CPA 65/ Production Manager/ Photographer/ McKnight Fellow/ Generals and Admirals/ Physicist 67
48	Accountant/ Economist/ Carpenter/ Salesmen, Encyclopedia/ Sales Manager/ Pulitzer Prize/ Judge/ Lawyer 49/ Newsmen/ Banker 64/ Physicist 27/ NIAL Member
47	Salesmen, Real Estate/ Engineer/ Corp Pres 35/ Purch Agent/ Banker 34/ Lawyer 27/ Corp Pres 65/ Artist
46	Architect/ Political Scientist/ Advertising Men/ Mathematician/ Author/ Interior Decorator
45	Farmer 67
44	
43	
42	
41	
40	
39	
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

Predictive Validity

The predictive validity of these scales can be demonstrated by reanalyzing some SVIBs collected by Berdie (1960) and Schletzer (1963). Berdie identified students who graduated from the University of Minnesota in curricula that are highly predictive of eventual occupations, specifically, accounting, dentistry, journalism (which is less predictive than the others), law, mechanical engineering, and medicine. Each of these students had completed the SVIB as high school seniors, and

TABLE 43

MALE OCCUPATIONS MEAN SCORES ON THE MERCHANDISING SCALE

M standard score	Sample
70	
69	
68	
67	
66	
65	
64	
63	
62	Salesmen, 3M
61	Salesmen, 3M applicants/ Bus Ed Teacher
60	
59	Credit Manager
58	Cham Comm Exec/ Salesmen, PG&E/ Salesmen, Encyclopedia
57	Salesmen, Steel/ Office Worker/ Salesmen, Computer
56	Pharmacist/ YMCA Sec/ Funeral Director/ Salesmen, Auto/ Purch Agent
55	Comm Rec Director
54	Rehab Counselor/ Accountant/ Soc Sci Teacher/ Guidance Counselor/ CPA 65/ Sales Manager/ YMCA Staff/ Interior Decorator
53	Salesmen, Real Estate/ Air Force Officer/ Legislator/ School Super 65
52	Salesmen, Life 31/ Personnel Director/ Elem Teacher/ Army Officer/ CPA 44/ Navy Officer/ Banker 64/ Salesmen, Life 66
51	YMCA PD/ Corp Pres 35/ Optometrist/ Banker 34/ Petroleum Engineer/ Math-Science Teacher
50	Vet 66/ Public Admin/ Social Worker 53/ Advertising Men
49	Production Manager/ Librarian/ Computer Programmer/ Generals and Admirals/ Physical Therapist/ Music Teacher
48	Priest/ School Super 30/ Printer/ Urologist/ Policemen/ Lawyer 49/ Minister 65/ Corp Pres 65/ Physiatrist/ Student Personnel/ Governor
47	Radiologist/ Carpenter/ Farmer 36/ Pilot /Minister 27/ Animal Husbandry Prof/ Social Worker 67/ Farmer 67/ Osteopath/ Orthopedic Surgeon/ Astronaut/ Vet 49
46	Psychologist 47/ Pediatrician/ Unitarian Minister/ Forest Service Men/ Musician
45	Sociologist/ Chemist/ Economist/ Dentist/ Psychiatrist/ Psychologist 49/ Neurological Surgeon
44	Judge/ Psychologist 31/ Engineer/ Surgeon/ Mpls Symphony/ Lawyer 27/ Newsmen/ Physician 49/ Internist
43	Architect/ Biologist
42	Physician 27/ Psychologist (Exp)/ McKnight Fellow/ Political Scientist
41	Photographer/ Pathologist/ Mathematician
40	Danforth Fellow/ Physicist 67
39	Physicist 27/ Authors/ Anthropologist
38	Pulitzer Prize/ Artist
37	
36	NIAL Member
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

Berdie's report indicates that there were substantial relationships between their high school SVIB profile and the curriculum they selected. Schletzer, for her doctoral dissertation, located and retested these same students approximately 4 yr. after they graduated from the University of Minnesota, roughly 8–10 yr. after the initial testing. Although these groups were small, approximately 30 in each of the six curricula, the results are very important as no other occupational samples have been tested before college and retested several years later after entry into the occupation.

The relevant data are presented in Tables 57 and 58. In Table 57 are listed, for the total group, test and retest means, standard deviations, and test-retest correlations. For each of the occupational subgroups, the test and retest

TABLE 44

MALE OCCUPATIONS MEAN SCORES ON THE MILITARY SCALE

M standard score	Sample
70	
69	
68	
67	
66	
65	
64	Army Officer/ Generals and Admirals
63	
62	
61	Policemen
60	Air Force Officer
59	
58	YMCA PD/ Pilot
57	
56	Navy Officer/ Astronaut/ Governor
55	Comm Rec Director
54	Salesmen, 3M applicants/ Rehab Counselor/ Public Admin/ Credit Manager/ Office Worker/ Salesmen, PG&E/ YMCA Staff/ Legislator/ Salesmen, Computer/ Physical Therapist
53	Salesmen, 3M/ Vet 66/ Cham Comm Exec/ Engineer/ Personnel Director/ Salesmen, Auto/ Optometrist/ CPA 44/ Farmer 67/ Osteopath/ Vet 49
52	Accountant/ Soc Sci Teacher/ Salesmen, Steel/ Production Manager/ Farmer 36/ Purch Agent/ Funeral Director/ Lawyer 27/ School Super 65/ Banker 64/ Bus Ed Teacher/ Petroleum Engineer/ Math-Science Teacher
51	Salesmen, Real Estate/ Pharmacist/ Salesmen, Life 31/ Guidance Counselor/ Carpenter/ Salesmen, Encyclopedia/ Elem Teacher/ Forest Service Men/ Sales Manager/ Dentist/ Animal Husbandry Prof/ Orthopedic Surgeon/ Advertising Men
50	Architect/ YMCA Sec/ Physician 27/ School Super 30/ Printer/ CPA 65/ Urologist/ Banker 34/ Lawyer 49/ Physician 49/ Physiatrist/ Salesmen, Life 66/ Music Teacher/ Student Personnel
49	Judge/ Priest/ Social Worker 53/ Chemist/ Surgeon/ Computer Programmer/ Corp Pres 35/ Newsmen/ Psychiatrist/ Social Worker 67/ Neurological Surgeon
48	Psychologist 31/ Radiologist/ Biologist/ Psychologist (Exp)/ Minister 27/ Pathologist/ Authors
47	Psychologist 47/ Librarian/ Minister 65/ Physicist 27/ Psychologist 49/ Musician/ Internist/ Interior Decorator
46	Pediatrician/ Unitarian Minister/ Economist/ Photographer/ Political Scientist/ Pulitzer Prize/ Artist/ Mathematician
45	Sociologist/ Corp Pres 65/ Physicist 67/ Anthropologist
44	Mpls Symphony/ Danforth Fellow/ NIAL Member
43	
42	McKnight Fellow
41	
40	
39	
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

means are presented, and the mean changes larger than 4 points are identified. Four points, which is four-tenths of a standard deviation, is probably the smallest practical difference worth attending to among these small samples.

As an aid to interpretation of the data in Table 57, Table 58 was prepared. In this table are listed for each subgroup:

1. Their three highest scores as high school seniors.

2. Their three highest scores on retest, when

TABLE 45

MALE OCCUPATIONS MEAN SCORES ON THE MUSIC SCALE

M standard score	Sample
70	
69	
68	
67	Music Teacher
66	
65	
64	McKnight Fellow/ Musician/ Mpls Symphony
63	
62	
61	Danforth Fellow
60	NIAL Member
59	
58	Librarian/ Minister 65
57	Unitarian Minister/ Minister 27/ Physiatrist
56	Psychologist 47/ Pediatrician/ Economist/ Photographer/ Political Scientist/ Pulitzer Prize
55	Priest/ Sociologist/ Social Worker 53/ Elem Teacher/ Newsmen/ Psychiatrist/ Interior Decorator/ Anthropologist
54	YMCA Sec/ Psychologist 31/ Biologist/ Social Worker 67/ Artist/ Pathologist/ Internist
53	Radiologist/ Psychologist (Exp)/ Computer Programmer/ Physical Therapist/ Psychologist 49/ Physicist 67/ Orthopedic Surgeon/ Neurological Surgeon/ Student Personnel
52	Salesmen, 3M applicants/ Cham Comm Exec/ Architect/ Soc Sci Teacher/ YMCA PD/ Credit Manager/ Printer/ Urologist/ Surgeon/ Optometrist/ Comm Rec Director/ Salesmen, Computer/ School Super 65/ Physician 49/ Bus Ed Teacher/ Mathematician/ Math-Science Teacher/ Advertising Men
51	Public Admin/ Accountant/ Chemist/ Office Worker/ Guidance Counselor/ Personnel Director/ Salesmen, Encyclopedia/ YMCA Staff/ Legislator/ CPA 44/ Physicist 27/ Salesmen, Life 66/ Osteopath
50	Rehab Counselor/ Salesmen, 3M/ Pharmacist/ Salesmen, Steel/ Salesmen, PG&E/ School Super 30/ CPA 65/ Carpenter/ Salesmen, Auto/ Army Officer/ Generals and Admirals/ Judge/ Navy Officer/ Corp Pres 65/ Animal Husbandry Prof/ Astronaut/ Authors
49	Physician 27/ Salesmen, Life 31/ Engineer/ Production Manager/ Policemen/ Pilot/ Dentist/ Funeral Director/ Air Force Officer/ Banker 64
48	Vet 66/ Farmer 36/ Purch Agent/ Forest Service Men/ Sales Manager/ Banker 34/ Lawyer 49/ Corp Pres 35/ Lawyer 27/ Petroleum Engineer/ Vet 49
47	Salesmen, Real Estate/ Governor
46	
45	
44	Farmer 67
43	
42	
41	
40	
39	
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

they were roughly 3–4 yr. beyond their final degree.

3. The three largest gains for each sample, which frequently were not among the highest scores on either testing.

Although, again, there is no one index to determine the predictive efficiency of these scales, the results are reassuringly meaningful. The scales that each group scored highest on, both at test and retest, are, with a few exceptions,

TABLE 46

MALE OCCUPATIONS MEAN SCORES ON THE NATURE SCALE

M standard score	Sample
70	
69	
68	
67	
66	
65	
64	
63	
62	
61	
60	
59	Farmer 36/ Animal Husbandry Prof
58	Forest Service Men
57	Vet 49
56	Vet 66/ Biologist
55	Orthopedic Surgeon
54	Radiologist/ Urologist/ Physiatrist
53	Pediatrician/ Surgeon/ Minister 27/ Physical Therapist/ Farmer 67/ Osteopath/ Pathologist/ Neurological Surgeon
52	Public Admin/ YMCA Sec/ Dentist/ Physician 49/ Psychiatrist/ Artist/ Internist/ NIAL Nember
51	YMCA PD/ Physician 27/ Comm Rec Director/ Generals and Admirals/ Math-Science Teacher
50	Rehab Counselor/ Architect/ Chemist/ School Super 30/ Unitarian Minister/ Guidance Counselor/ Mpls Symphony/ School Super 65/ Anthropologist/ Governor
49	Psychologist 47/ Soc Sci Teacher/ Social Worker 53/ Psychologist 31/ Credit Manager/ Psychologist (Exp)/ Personnel Director/ Carpenter/ Elem Teacher/ Policemen/ Librarian/ Pulitzer Prize/ Army Officer/ Banker 34/ Legislator/ CPA 44/ Psychologist 49/ Music Teacher/ Interior Decorator
48	Pharmacist/ Sociologist/ Accountant/ Engineer/ McKnight Fellow/ Pilot/ Optometrist/ Lawyer 49/ Corp Pres 35/ Funeral Director/ Minister 65/ Physicist 27/ Bus Ed Teacher/ Musician/ Physicist 67/ Astronaut/ Mathematician/ Authors/ Student Personnel
47	Priest/ Office Worker/ Salesmen, Life 31/ Salesmen, PG&E/ Printer/ Production Manager/ Photographer/ Sales, Encyclopedia/ Judge/ YMCA Staff/ Computer Programmer/ Corp Pres 65/ Banker 64/ Social Worker 67/ Petroleum Engineer
46	Salesmen, Real Estate/ Economist/ Sales Manager/ Danforth Fellow/ Navy Officer
45	Cham Comm Exec/ Salesmen, Steel/ CPA 65/ Purch Agent/ Political Scientist/ Air Force Officer/ Lawyer 27/ Newsmen/ Advertising Men
44	Salesmen, Life 66
43	Salesmen, Auto/ Salesmen, Computer
42	Salesmen, 3M
41	Salesmen, 3M applicants
40	
39	
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

those most related to their work. The exceptions were almost all due to the Adventure scale; this was among the three highest scores for five of the samples at test, and remained among the highest for three at the retest. This personality scale just does not operate as do the vocational interests scales. Other than this scale, the other high scores were appropriate. The accountants scored highest on the business-oriented scales: Office Practices, Sales, and Merchandising; the dentists and physicians scored highest on the Medical Service and Science scales, and so forth.

The patterns of high scores on both the test

TABLE 47

MALE OCCUPATIONS MEAN SCORES ON THE OFFICE PRACTICES SCALE

M standard score	Sample
70	
69	
68	
67	
66	
65	Bus Ed Teacher
64	
63	
62	
61	Baker 34/ Banker 64
60	
59	Office Worker
58	Accountant/ Soc Sci Teacher/ Credit Manager
57	
56	
55	Rehab Counselor/ Guidance Counselor/ Comm Rec Director/ Math-Science Teacher
54	Salesmen, 3M/ YMCA Sec/ Salesmen, PG&E/ CPA 65/ CPA 44/ School Super 65
53	Public Admin/ Purch Agent/ Policemen/ Legislator/ Funeral Director
52	Salesmen, 3M applicants/ Pharmacist/ Farmer 67/ Governor
51	Vet 66/ Cham Comm Exec/ YMCA PD/ School Super 30/ Printer/ Carpenter/ Elem Teacher/ Salesmen, Auto/ Librarian/ Army Officer/ YMCA Staff/ Minister 27/ Air Force Officer
50	Salesmen, Real Estate/ Social Worker 53/ Salesmen, Steel/ Personnel Director/ Farmer 36/ Forest Service Men/ Optometrist/ Navy Officer/ Minister 65/ Music Teacher
49	Priest/ Salesmen, Encyclopedia/ Sales Manager/ Computer Programmer/ Lawyer 49/ Generals and Admirals/ Physical Therapist/ Petroleum Engineer/ Vet 49
48	Salesmen, Life 31/ Production Manager/ Salesmen, Computer/ Corp Pres 35/ Physiatrist/ Animal Husbandry Prof/ Osteopath/ Interior Decorator/ Student Personnel
47	Urologist/ Social Worker 67
46	Psychologist 47/ Sociologist/ Radiologist/ Pediatrician/ Unitarian Minister/ Pilot/ Dentist/ Mathematician
45	Judge/ Psychologist 31/ Chemist/ Lawyer 27/ Corp Pres 65/ Psychologist 49/ Musician/ Orthopedic Surgeon
44	Biologist/ Economist/ Political Scientist/ Mpls Symphony/ Physician 49/ Psychiatrist/ Advertising Men
43	Engineer/ Psychologist (Exp)/ Newsmen/ Physicist 27/ Salesmen, Life 66/ Internist/ Astronaut/ Neurological Surgeon
42	Physician 27/ Surgeon/ Pathologist
41	Architect/ McKnight Fellow/ Pulitzer Prize/ Physicist 67
40	Danforth Fellow/ Authors/ Anthropologist
39	Photographer/ NIAL Member
38	
37	Artist
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

and retest administrations again make it very clear that the scales have a substantial relationship to the individual's occupational choice.

The gains reported in Table 58 are worthy of note also as they indicate areas of interests showing the greatest increase over the college years, irrespective of initial level of score. In general, the changes should please university faculties. The largest gain for all six groups was on the Teaching scale, ranging from 7 points among the lawyers to 10 points among the dentists, a full standard deviation. All of the groups also showed increased scores on the Art,

Music, and Writing scales, areas particularly important to a liberal arts college faculty. Interestingly, this tendency was less pronounced among the accountants who were primarily students drawn from the Business School. However, the other non-liberal-arts sample—the mechanical engineers—showed the same increases as the other students.

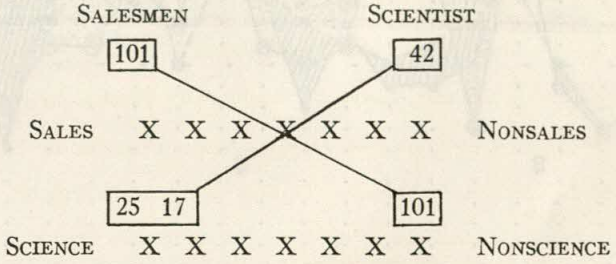
These increases are very similar to those found among the Stanford students in the reliability study reported earlier, showing that interests in teaching and in cultural activities increase during the college years. The same pattern, though less pronounced, appeared in the Harvard test-retest profile in Table 32. These highly selected students had higher scores in these areas when initially tested, and their scores did not increase as much as the Minnesota and Stanford students.

Salsemen and Scientists

Further predictive validity information has been developed by studying two subsets of individuals in the norm group. As these men have been tested twice, once in 1930 as 16-yr.-olds and again in 1966 as 52-yr.-olds, it has been possible to study how their test results relate to their adult occupations.

Of the 1,943 boys tested in 1930, 1,214 (63%) have been retested in 1966–67. Twenty-five percent could not be located currently, 4% refused to cooperate in this study, and 8% were deceased. (When these basic scales were normed, only 647 retests were available.)

When retested as adults, these men were asked to fill in a short checklist describing their jobs. This checklist contained "semantic differential" type items, that is, pairs of statements reflecting opposite extremes, and each man was asked to indicate where his job fell on a line between these pairs. Two of the items were "sales versus nonsales" and "science versus nonscience." From the total sample, two subgroups of individuals were identified, using their answers to these two items. The first included 101 men who marked the opposing extremes, "sales, nonscience," as descriptive of their jobs; the other sample included 42 men who marked the other extremes, "science, nonsales." The following diagram shows graphically how these groups were selected and reports the number in each group:



The test (age 16) and retest (age 52) SVIB basic scale profiles for these two groups are presented in Fig. 3.

As the mean profiles indicate, the Sales and Science scales separated these groups fairly well at age 16, and by age 52 the difference had increased. At age 16, the differences were roughly 1 *SD*, increasing to about 1.5 or 2 *SD*s between the adults. The larger differences between the adult samples were created because the salesmen increased more on the sales-oriented scales, the scientists on the science-oriented scales, not because either of them had decreasing scores. Whether these increases came before or after the occupational experience cannot be determined from these data; the latter seems more plausible.

These mean differences on the basic scales, though large and statistically far beyond the chance level, were not as large as the differences between these two samples on the regular SVIB occupational scales. For example, on the SVIB Chemist scale, the samples had the following test and retest means—test: Scientist $M = 39$, Salesman $M = 26$, $D_M = 13$; retest: Scientist $M = 43$, Salesman $M = 17$, $D_M = 26$. On the SVIB Sales Manager scale, the samples had the following test and retest means—test:

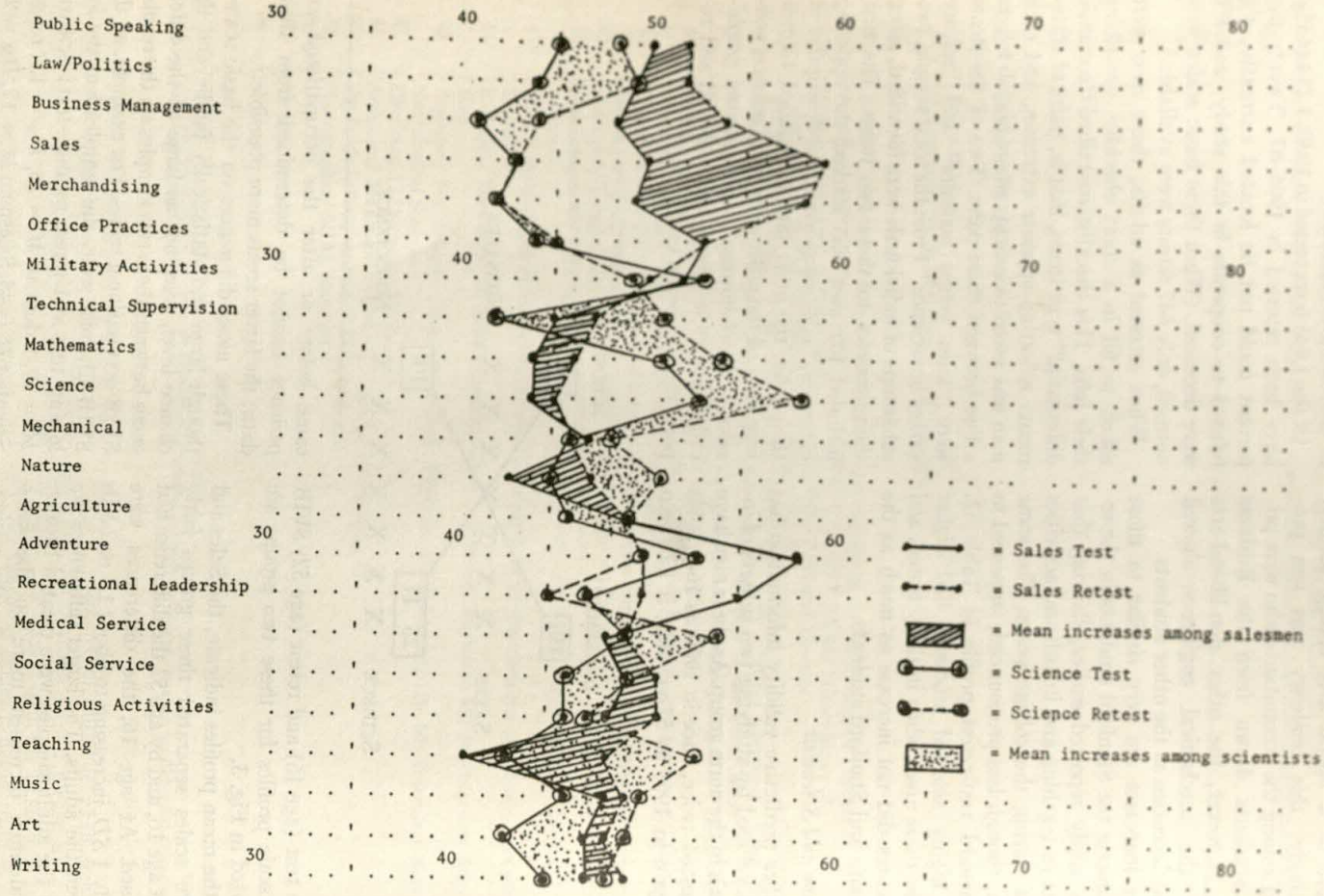


FIG. 3. Test and retest profiles for salesmen and scientist samples.

TABLE 48

MALE OCCUPATIONS MEAN SCORES ON PUBLIC SPEAKING SCALE

M standard score	Sample
70	
69	
68	
67	Legislator
66	Governor
65	Unitarian Minister/ Minister 65
64	
63	Priest/ Cham Comm Exec/ Salesmen, Real Estate/ Political Scientist/ Salesmen, Computer
62	Salesmen, 3M applicants/ Minister 27/ School Super 65/ Salesmen, Life 66
61	Salesmen, 3M/ Comm Rec Director/ Danforth Fellow/ Social Worker 67
60	YMCA Sec/ Social Worker 53/ YMCA Staff/ Judge/ Newsmen/ Student Personnel
59	Sociologist/ Economist/ Elem Teacher/ Lawyer 49
58	Soc Sci Teacher/ Salesmen, Steel/ Guidance Counselor/ Generals and Admirals/ Navy Officer/ Astronaut
57	Rehab Counselor/ Credit Manager/School Super 30/Air Force Officer/McKnight Fellow/ Lawyer 27
56	Public Admin/ CPA 65/ Optometrist/ Corp Pres 65/ Music Teacher/ Bus Ed Teacher
55	Salesmen, Life 31/ Librarian/ Pulitzer Prize/ Army Officer/ Physiatrist/ Animal Husbandry Prof
54	Psychologist 47/ YMCA PD/ Salesmen, PG&E/ Personnel Director/ Computer Programmer/ Physical Therapist/ Psychologist 49/ Petroleum Engineer/ Neurological Surgeon
53	Vet 66/ Pediatrician/ Sales Manager/ Psychiatrist/ Orthopedic Surgeon/ Anthropologist
52	Psychologist 31/ Funeral Director/ Biologist/ Urologist/ Salesmen, Auto/ Physicist 67/ Internist/ Math-Science Teacher/ NIAL Member
51	Salesmen, Real Estate/ Accountant/ Radiologist/ Photographer/ Psychologist (Exp)/ Surgeon/ Policemen/ CPA 44/ Banker 64/ Osteopath/ Authors/ Interior Decorator/ Advertising Men
50	Office Worker/ Printer/ Production Manager/ Purch Agent/ Mpls Symphony/ Banker 34/ Physician 49/ Musician/Pathologist
49	Pharmacist/ Farmer 36/ Pilot/ Vet 49
48	Physician 27/ Corp Pres 35/ Forest Service Men/ Dentist/ Mathematician
47	Chemist/ Engineer/ Physicist 27
46	Architect/ Carpenter
45	Farmer 67
44	Artist
43	
42	
41	
40	
39	
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

Salesman $M = 35$, Scientist $M = 27$, $D_M = 8$; retest: Salesman $M = 41$, Scientist $M = 20$, $D_M = 21$.

These larger differences, exceeding 2 *SD*s among the adult samples, can probably be traced to the relative heterogeneity of the occupational scales compared to the basic scales. Each of these two occupational scales, Chemist and Sales Manager, contains both science and sales components. In the Chemist scale, these components are weighted positively and negatively, respectively; in the Sales Manager scale, the reverse is true. Thus the occupational scales take advantage of both components for maximum separation, while the basic scales sacrifice some of this differentiation for easier interpretability. Note again that there were differential increases (and,

TABLE 49

MALE OCCUPATIONS MEAN SCORES ON THE RECREATIONAL LEADERSHIP SCALE

M standard score	Sample
70	
69	
68	
67	
66	
65	
64	
63	
62	
61	YMCA PD
60	
59	Salesmen, 3M applicant/ Comm Rec Director
58	Salesmen, 3M/ YMCA Staff/ Physical Therapist
57	
56	
55	Policemen
54	Salesmen, Steel/ Elem Teacher/ Salesmen, Auto/ Salesmen, Computer
53	Soc Sci Teacher/ Salesmen, Life 66/ Social Worker 67
52	Credit Manager/ CPA 65/ Pilot/ Army Officer/ Legislator/ Air Force Officer/ Math-Science Teacher
51	YMCA Sec/ Salesmen, PG&E/ Forest Service Men/ School Super 65/ Osteopath/ Petroleum Engineer/ Orthopedic Surgeon/ Astronaut/ Vet 49
50	Vet 66/ Cham Comm Exec/ Printer/ Guidance Counselor/ Urologist/ Computer Programmer/ Generals and Admirals/ CPA 44/Newsmen/ Bus Ed Teacher
49	Rehab Counselor/ Social Worker 53/ Office Worker/ Salesmen, Life 31/ Personnel Director/ Salesmen, Encyclopedia/ Optometrist/ Dentist/ Navy Officer/ Minister 65/ Banker 64/ Physiatrist/ Neurological Surgeon
48	Priest/ Salesmen, Real Estate/ Pharmacist/ Public Admin/ Accountant/ Radiologist/ Pediatrician/ Production Manager/ Carpenter/ Farmer 36/ Surgeon/ Purch Agent/ Sales Manager/ Animal Husbandry Prof/ Farmer 67/ Student Personnel/ Governor
47	School Super 30/ Unitarian Minister/ Lawyer 49/ Minister 27/ Funeral Director/ Lawyer 27/ Corp Pres 65/ Physician 49/ Psychiatrist/ Musician
46	Psychologist 47/ Physician 27/ Chemist/ Banker 34/ Advertising Men
45	Sociologist/ Engineer/ Danforth Fellow/ Corp Pres 35/ Music Teacher/ Internist
44	Psychologist 31/ Photographer/ Mpls Symphony/ Judge/ Pathologist
43	Architect/ Biologist/ Economist/ McKnight Fellow/ Pulitzer Prize/ Psychologist 49/ Artist/ Authors
42	Psychologist (Exp)/ Political Scientist/ Librarian
41	Physicist 67/ Anthropologist
40	Physicist 27
39	Mathematician/ Interior Decorator
38	NIAL Member
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

contrary to the findings on the basic scales, actual decreases on the nonappropriate scale) on these scales for these two differing samples.

Construct Validity, Interest Constellations, and Implications

All three types of validity information reported here, content, concurrent, and predictive, support the conclusion that these scores are related to occupational choice. Collectively, they also argue for the existence of the fourth type of validity that psychologists have found useful: construct validity. Construct validity refers to the concepts that underlie test behavior, concepts which may not be directly accessible for measurement but which provide

TABLE 50

MALE OCCUPATIONS MEAN SCORES ON RELIGIOUS ACTIVITIES SCALE

M standard score	Sample
75	
74	
73	
72	
71	YMCA Sec/ Minister 27
70	Minister 65
69	YMCA Staff
68	
67	Priest
66	YMCA PD
65	
64	Unitarian Minister
63	
62	
61	Comm Rec Director/ Legislator
60	Rehab Counselor/ Phsyical Therapist
59	Salesmen, 3M/ Cham Comm Exec/ Social Worker 53/ Guidance Counselor/ School Super 65/ Student Personnel
58	Salesmen, 3M applicant/ Credit Manager/ Salesmen, Encyclopedia/ Elem Teacher/ Danforth Fellow/ Salesmen, Life 66/ Social Worker 67/ Music Teacher
57	Soc Sci Teacher
56	Salesmen, Steel/ Optometrist/ Banker 64/ Physiatrist/ Bus Ed Teacher
55	Sociologist/ School Super 30/ CPA 65/ Funeral Director/ Farmer 67
54	Vet 66/ Librarian/ Salesmen, Computer/ Math-Science Teacher/ Governor
53	Computer Programmer/ Judge/ Lawyer 49/ Air Force Officer/ Animal Husbandry Prof/ Petroleum Engineer
52	Salesmen, Life 31/ Policemen/ Mpls Symphony/ Navy Officer/ Astronaut/ Vet 49
51	Public Admin/ Radiologist/ Office Worker/ Personnel Director/ Salesmen, Auto/ Banker 34/ Political Scientist/ Newsmen/ Generals and Admirals/ Orthopedic Surgeon/ Neurological Surgeon
50	Pharmacist/ Accountant/ Pediatrician/ Biologist/ Urologist/ Economist/ Pulitzer Prize/ Army Officer/ CPA 44/ Corp Pres 65/ Musician/ Osteopath
49	Psychologist 47/ Salesmen, PG&E/ Printer/ Production Manager/ Farmer 36/ McKnight Fellow/ Sales Manager/ Physician 49/ Interior Decorator
48	Carpenter/ Purch Agent/ Corp Pres 35/ Psychologist 49/ Mathematician/ NIAL Member
47	Salesmen, Real Estate/ Engineer/ Photographer/ Surgeon/ Forest Service Men/ Dentist/ Lawyer 27/ Physicist 27/ Psychiatrist
46	Pilot/ Physicist 67/ Pathologist/ Internist
45	Architect/ Physician 27/ Chemist/ Authors/ Anthropologist/ Advertising Men
44	Psychologist 31/ Psychologist (Exp)
43	Artist
42	
41	
40	
39	
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

some explanatory power for the parallel features of test responses and actual behavior. What follows here is a brief attempt to make psychological sense of these clusters of interests and to suggest a mechanism to explain an individual's answers to these items.

The underlying concepts here are the related constellations of interests. Somehow these con-

TABLE 51

MALE OCCUPATIONS MEAN SCORES ON THE SALES SCALE

M standard score	Sample
70	
69	
68	
67	Salesmen, 3M applicants
66	Salesmen, 3M
65	
64	Salesmen, PG&E/ Salesmen, Encyclopedia
63	Salesmen, Auto/ Salesmen, Computer
62	
61	Salesmen, Steel/ Salesmen, Life 31/ Salesmen, Life 66
60	
59	Salesmen, Real Estate/ Sales Manager/ Bus Ed Teacher
58	
57	
56	YMCA Sec/ Credit Manager/ Legislator
55	Pharmacist/ Cham Comm Exec/ Funeral Director
54	Office Worker/ Guidance Counselor
53	Rehab Counselor/ Soc Sci Teacher/ YMCA Staff/ School Super 65
52	YMCA PD/ Corp Pres 35/ Air Force Officer/ Purch Agent/ Comm Rec Director/ Banker 34/ Advertising Men/ Banker 64/ Farmer 67
51	Vet 66/ Corp Pres 65/ Petroleum Engineer/ Vet 49
50	Accountant/ Social Worker 53/ CPA 65/ Production Manager/ Personnel Director/ Carpenter/ Optometrist/ Physical Therapist/ Music Teacher/ Interior Decorator/ Math-Science Teacher/ Governor
49	School Super 30/ Farmer 36/ Elem Teacher/ Policemen/ Minister 27/ Navy Officer/ Minister 65
48	Priest/ Printer/ Army Officer/ CPA 44/ Physiatrist/ Social Worker 67/ Osteopath
47	Public Admin/ Radiologist/ Unitarian Minister/ Urologist/ Forest Service Men/ Pilot/ Computer Programmer/ Lawyer 49/ Animal Husbandry Prof/ Musician/ Orthopedic Surgeon
46	Psychologist 47/ Dentist/ Generals and Admirals/ Astronaut/ Neurological Surgeon/ Student Personnel
45	Pediatrician/ Chemist/ Engineer/ Surgeon/ Librarian/ Physician 49/ Psychiatrist/ Psychologist 49
44	Sociologist/ Mpls Symphony/ Judge/ Newsmen/ Lawyer 27/ Internist
43	Architect/ Psychologist 31/ Physician 27/ Biologist/ Economist/ Photographer
42	Psychologist (Exp)/ McKnight Fellow/ Political Scientist/ Danforth Fellow/ Pathologist/ Mathematician
41	Pulitzer Prize/ Physicist 27/ Physicist 67/ Artist/ Authors
40	Anthropologist
39	NIAL Member
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

stellations dictate, at least to some degree, the individual's feeling toward activities which fall within the bounds of the constellation. It might be fruitful in future research to consider an individual's answer to a specific item (and thus his feeling toward a specific behavior) as a combination of his perception of which con stellation an item belongs to and of his attraction toward that constellation generally. Thus, the two salient variables are his evaluation of constellation membership for an item and his attraction for the activities within the constellation—in a convenient shorthand, Perception × Attraction.

A clarifying example: A salesman would answer Like to the item "Sell Fuller brushes door-to-door" because he would perceive that as a sales item and he likes sales activities; a scientist would mark it Dislike because he also perceives it as a sales item and dislikes those activities. But if confronted with the item, "Work for the National Academy of Sciences,

TABLE 52

MALE OCCUPATIONS MEAN SCORES ON THE SCIENCE SCALE

M standard score	Sample
70	
69	
68	
67	
66	
65	
64	
63	
62	Physicist 27/ Physicist 67
61	Chemist
60	Biologist/ Psychologist (Exp)/ Pathologist
59	Psychologist 31/ Animal Husbandry Professor/ Neurological Surgeon
58	Pediatrician/ Urologist/ Internist/ Mathematician/ Math-Science Teacher
57	Radiologist/ Engineer/ Air Force Officer/ Physiatrist/ Psychiatrist/ Psychologist 49/ Orthopedic Surgeon/ Astronaut
56	Psychologist 47/ Surgeon/ Anthropologist
55	Physician 27/ Economist/ Optometrist/ Computer Programmer/ Danforth Fellow/ Physician 49/ Physical Therapist/ Petroleum Engineer
54	Army Officer
53	Pilot/ Dentist/ Generals and Admirals/ Osteopath
52	Vet 66/ Pharmacist/ Public Admin/ Architect/ Sociologist/ School Super 30/ Mpls Symphony/ Salesmen, Computer/ CPA 44/ Navy Officer
51	Rehab Counselor/ YMCA PD/ Unitarian Minister/ Production Manager/ Personnel Director/ Elem Teacher/ Librarian/ Minister 27/ NIAL Member/ Student Personnel
50	Accountant/ Guidance Counselor/ Corp Pres 35/ Farmer 36/ Forest Service Men/ Political Scientist/ Policemen/ School Super 65/ Corp Pres 65/ Musician/ Artist/ Vet 49
49	YMCA Sec/ Social Worker 53/ Credit Manager/ Salesmen, PG&E/ Printer/ Photographer/ Judge/ Laywer 49
48	Salesmen, 3M/ Priest/ Salesmen, Steel/ Office Worker/ CPA 65/ Purch Agent/ Pulitzer Prize/ Comm Rec Director/ Lawyer 27/ Music Teacher/ Bus Ed Teacher
47	Soc Science Teacher/ Carpenter/ McKnight Fellow/ Sales Manager/ YMCA Staff/ Legislator/ Minister 65/ Social Worker 67/ Governor
46	Salesmen, 3M applicant/ Cham Comm Exec/ Advertising Men/ Banker 64/ Author
45	Salesmen, Encyclopedia/ Salesmen, Auto/ Newsmen
44	Salesmen, Life 31/ Funeral Director/ Banker 34
43	Salesmen, Real Estate
42	Salesmen, Life 66/ Farmer 67/ Interior Decorator
41	
40	
39	
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

appearing before Congress, private foundations, and other prestige organizations, selling the idea that science must be broadly supported by our society," they both might answer Like as one would perceive it as a sales activity, the other as science. In this illustration, their perceptions might be modified substantially by substituting other verbs for "selling," such as persuading or, better yet, educating or informing.

This formulation of organizing constellations as the constructs underlying interest measurement is undoubtedly too simple, and only partially correct. Yet it does provide a way to begin to understand more than the empirical facts, to lead us out of the wilderness of means

TABLE 53

MALE OCCUPATIONS MEAN SCORES ON THE SOCIAL SERVICE SCALE

M standard score	Sample
70	
69	YMCA Sec
68	
67	Social Worker 53/ YMCA Staff/ Minister 65
66	Social Worker 67
65	YMCA PD/ Minister 27
64	
63	Rehab Counselor/ Priest/ Unitarian Minister/ Guidance Counselor
62	Soc Sci Teacher
61	Psychologist 47/ Comm Rec Director/ Student Personnel
60	Salesmen, 3M/ Legislator/ Physical Therapist
59	Salesmen, 3M applicants/ Sociologist/ Elem Teacher/ Danforth Fellow/ School Super 65/ Music Teacher
58	Salesmen, Encyclopedia/ Bus Ed Techer
57	School Super 30
56	Cham Comm Exec/ Credit Manager/ Librarian/ Physiatrist/ Salesmen, Life 66/ Math-Science Teacher/ Governor
55	Pediatrician/ Salesmen, Life 31/ Policemen/ Political Scientist/ Mpls Symphony/ Funeral Director/ Psychiatrist
54	Pharmacist/ Public Admin/ Salesmen, Steel/ Personnel Director/ Optometrist/ Lawyer 49/ Psychologist 49/ Musician/ Vet 49
53	Vet 66/ Salesmen, PG&E/ CPA 65/ Economist/ McKnight Fellow/ Newsmen/ Banker 64/ Farmer 67/ Osteopath
52	Judge/ Office Worker/ Printer/ Carpenter/ Salesmen, Auto/ Banker 34/ Salesmen, Computer/ CPA 44/ Internist
51	Accountant/ Radiologist/ Urologist/ Photographer/ Forest Service Men/ Computer Programmer/ Corp Pres 35/ Air Force Officer/ Physician 49/ Orthopedic Surgeon/ Interior Decorator/ NIAL Member
50	Salesmen, Real Estate/ Production Manager/ Psychologist (Exp)/ Farmer 36/ Sales Manager/ Pulitzer Prize/ Lawyer 27/ Corp Pres 65/ Anthropologist/ Neurological Surgeon
49	Psychologist 31/ Physician 27/ Biologist/ Surgeon/ Purch Agent/ Dentist/ Navy Officer/ Animal Husbandry Prof/ Petroleum Engineer/ Pathologist/ Advertising Men
48	Army Officer/ Mathematician
47	Architect/ Chemist/ Engineer/ Pilot/ Generals and Admirals/ Physicist 27/ Physicist 67/ Astronaut/ Authors
46	Artist
45	
44	
43	
42	
41	
40	
39	
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

and correlations, and it has some immediate implications for the techniques of measurement:

1. If item perception is important, then item clarity is a paramount concern. Interindividual agreement as to what an item means should be very high. This can be achieved in several ways: by using short familiar items; by using longer very specific items; or by using powerful stereotypes such as occupational titles.

TABLE 54

MALE OCCUPATIONS MEAN SCORES ON THE TEACHING SCALE

M standard score	Sample
70	
69	
68	
67	
66	Music Teacher
65	
64	
63	Elem Teacher
62	Librarian/ Danforth Fellow/ Minister 65/ School Super 65
61	School Super 30/ Unitarian Minister/ Guidance Counselor/ Minister 27/ Bus Ed Teacher/ Student Personnel
60	Rehab Counselor/ Soc Sci Teacher/ Social Worker 67
59	Priest/ Sociologist/ Mpls Symphony/ Comm Rec Director/ YMCA Staff/ Physiatrist
58	Social Worker 53/ Biologist/ McKnight Fellow/ Political Scientist/ Mathematician/ Math-Science Teacher/ Neurological Surgeon
57	Psychologist 47/ Psychologist 31/ Pediatrician/ Legislator/ Physical Therapist/ Psychologist 49
56	YMCA Sec/ Economist/ Psychiatrist/ Animal Husbandry Prof/ Musician/ Physicist 67/ Pathologist/ Internist/ Anthropologist
55	YMCA PD/ Psychologist (Exp)/ Judge/ Salesmen, Encyclopedia/ Optometrist/ Pulitzer Prize/ Generals and Admirals/ Navy Officer/ Physicist 27/ Orthopedic Surgeon/ NIAL Member
54	Salesmen, 3M applicant/ Cham Comm Exec/ Credit Manager/ Air Force Officer/ Surgeon/ Computer Programmer/ Lawyer 49/ Salesmen, Computer/ Newsmen
53	Salesmen, 3M/ Radiologist/ Photographer/ Army Officer/ Salesmen, Life 66/ Astronaut
52	Vet 66/ CPA 65/ Urologist/ Personnel Director/ Corp Pres 65/ Physician 49/ Governor
51	Public Admin/ Chemist/ Petroleum Engineer/ Interior Decorator
50	Physician 27/ Salesmen, Steel/ CPA 44
49	Accountant/ Lawyer 27/ Banker 64/ Osteopath
48	Salesmen, PG&E/ Printer/ Policemen/ Dentist/ Vet 49
47	Pharmacist/ Architect/ Office Worker/ Salesmen, Life 31/ Production Manager/ Advertising Men/ Authors
46	Engineer/ Corp Pres 35/ Carpenter/ Farmer 36/ Salesmen, Auto/ Forest Service Men/ Pilot/ Artist
45	Sales Manager
44	Funeral Director/ Banker 34/ Farmer 67
43	Salesmen, Real Estate/ Purch Agent
42	
41	
40	
39	
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

2. It suggests that forced-choice items are too complicated. To ask a person to sort out his perceptions of three items in a triad, then to determine how he feels about each of these and compare those feelings is creating an unnecessarily complex task for him. Worse, it means the same pattern of answers could result from several different psychological viewpoints. While that is inevitable, no matter what method is used, it should be minimized. (Forced-choice items may have some advantages in other contexts—but they have yet to be demonstrated empirically.)

3. The Perception × Attraction theory emphasizes that, to this point in history, little attention has been paid to the *perception* portion in interest measurement—all of the emphasis has been placed on the *attraction* aspect

TABLE 55

MALE OCCUPATIONS MEAN SCORES ON TECHNICAL SUPERVISION SCALE

M standard score	Sample
70	
69	
68	
67	
66	
65	
64	
63	
62	
61	
60	
59	
58	
57	Air Force Officer/ Petroleum Engineer
56	Production Manager
55	Chemist/ Army Officer
54	Engineer/ Carpenter/ Pilot/ Salesmen, Computer
53	Personnel Director/ Navy Officer/ Orthopedic Surgeon/ Astronaut
52	Printer/ Neurological Surgeon
51	Public Admin/ Radiologist/ Salesmen, Steel/ Salesmen, PG&E/ Corp Pres 35/ Urologist/ Computer Programmer/ Corp Pres 65/ Math-Science Teacher
50	Salesmen, 3M applicant/ Salesmen, 3M/ Purch Agent/ Policemen/ Generals and Admirals/ CPA 44/ Physical Therapist/ Farmer 67
49	Vet 66/ Accountant/ Pediatrician/ Credit Manager/ Optometrist/ Comm Rec Director/ Legislator/ Physiatrist/ Bus Ed Teacher
48	Rehab Counselor/ Cham Comm/ Social Worker 53/ Office Worker/ CPA 65/ Psychologist (Exp)/ Surgeon/ Physicist 27/ Psychiatrist/ Psychologist 49/ Animal Husbandry Prof
47	YMCA Sec/ YMCA PD/ Psychologist 31/ Guidance Counselor/ Economist/ Farmer 36/ Elem Teacher/ Salesmen, Auto/ Sales Manager/ YMCA Staff/ Physicist 67/ Osteopath/ Pathologist
46	Psychologist 47/ Pharmacist/ Soc Sci Teacher/ Unitarian Minister/ Forest Service Men/ Dentist/ Lawyer 49/ Funeral Director/ School Super 65/ Banker 64/ Physician 49/ Social Worker 67/ Internist/ Vet 49
45	Architect/ Biologist/ Student Personnel
44	Priest/ Sociologist/ Physician 27/ Salesmen, Encyclopedia/ Librarian/ Banker 34/ Minister 27/ Minister 65/ Musician/ Mathematician
43	Salesmen, Real Estate/ School Super 30/ Political Scientist/ Mpls Symphony/ Anthropologist
42	Music Teacher/ Governor
41	Salesmen, Life 31/ Photographer/ Lawyer 27/ Advertising Men
40	Danforth Fellow/ Newsmen/ Salesmen, Life 66/ Interior Decorator
39	Judge/ Artist
38	McKnight Fellow/ NIAL Member
37	Pulitzer Prize/ Authors
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

by requiring the Like, Indifferent, or Dislike response. More should be done on perception, at least enough to determine if this way of thinking has any merit.

4. If the perception factor does prove useful, it might help explain the changes over time within a single individual as it would allow for two types of changes, one in perception, the other in actual attraction to the activities. Such a possibility is particularly appealing when studying teenage-to-adult samples as some of their differences may be traced to the teenagers' difficulty with the vocabulary. Perception certainly includes within it understanding; if a person does not understand an item, his perception will be affected. When

TABLE 56

MALE OCCUPATIONS MEAN SCORES ON THE WRITING SCALE

M standard score	Sample
70	
69	
68	
67	
66	
65	
64	Unitarian Minister/ Librarian/ Pulitzer Prize/ Newsmen
63	McKnight Fellow/ Authors
62	Political Scientist/ Danforth Fellow/ NIAL Member
61	Minister 65
60	Psychologist 47/ Judge/ Photographer/ Minister 27/ Sociologist
59	Anthropologist/ Neurological Surgeon/ Social Worker 63
58	Economist/ Elem Teacher/ Legislator/ Social Worker 67/ Music Teacher/ Student Personnel/ Advertising Men/ Cham Comm Exec/ Psychologist 31
57	Rehab Counselor/ Priest/ Mpls Symphony/ Lawyer 49/ School Super 65/ Physiatrist/ Psychiatrist/ Psychologist 49/ Governor/ YMCA Sec/ Pediatrician
56	School Super 30/ Psychologist (Exp)/ Salesmen, Encyclopedia/ Lawyer 27/ Musician
55	Salesmen, 3M applicants/ Salesmen, 3M/ Printer/ Guidance Counselor/ Biologist/ Personnel Director/ Comm Rec Director/ Army Officer/ Generals and Admirals/ Artist/ Internist/ Interior Decorator/ Public Admin/ Soc Sci Teacher
54	Credit Manager/ YMCA Staff/ Salesmen, Life 66/ Bus Ed Teacher/ Physicist 67/ Orthopedic Surgeon/ Radiologist
53	Urologist/ Surgeon/ Computer Programmer/ Salesmen, Computer/ Navy Officer/ Physical Therapist/ Pathologist/ YMCA PD
52	Salesmen, Steel/ Office Worker/ Salesmen, Life 31/ Salesmen, PG&E/ Air Force Officer/ Optometrist/ CPA 44/ Physician 49/ Astronaut/ Mathematician/ Architect
51	Corp Pres 35/ CPA 65/ Sales Manager/ Animal Husbandry Prof/ Accountant/ Chemist
50	Forest Service Men/ Policemen/ Pilot/ Corp Pres 65/ Osteopath/ Petroleum Engineer/ Math-Science Teacher/ Salesmen, Real Estate/ Physician 27
49	Engineer/ Production Manager/ Salesmen, Auto/ Dentist/ Banker 34/ Physicist 27/ Pharmacist
48	Vet 66/ Funeral Director/ Farmer 36/ Purch Agent/ Banker 64
47	Carpenter
46	Vet 49
45	
44	
43	Farmer 67
42	
41	
40	
39	
38	
37	
36	
35	

Note.—In some cases occupational groups were sampled and tested in different years (e.g., Banker, see Table 34). Thus numbers following samples indicate appropriate year.

understanding comes, his answer may change even though his attraction to that activity has not.

This brief theoretical side trip illustrates one possibility of further research in clustering interests, and in understanding the results. After 40 yr. of fact gathering, we must have some organizing theory, especially since the computer has made it possible to carry empiricism to ridiculous extremes.

USE OF THE BASIC SCALES

This paper has been almost solely concerned with the construction and evaluation of scales

TABLE 57

MEAN TEST AND RETEST PROFILES FOR THE BERDIE–SCHLETZER CURRICULAR GROUPS

Scale	Total sample						Accountants ($N = 24$)			Dentists ($N = 30$)			Journalists ($N = 21$)			Lawyers ($N = 32$)			Mechanical Engineers ($N = 38$)			Physicians ($N = 26$)		
	Test		Retest		r	C[a]	M		C	M		C	M		C	M		C	M		C	M		C
	M	SD	M	SD			Test	Retest		Test	Retest		Test	Retest		Test	Retest		Test	Retest		Test	Retest	
1 Public Speaking	53	10.8	57	11.0	.67	+4	52	55		46	47		55	61	+6	63	65		50	54	+4	54	59	+5
2 Law/Politics	53	10.6	57	10.6	.70	+4	55	60	+5	47	50		54	59	+5	61	66	+5	50	55	+5	52	55	
3 Business Management	49	10.3	50	10.5	.46		54	56		46	43		51	53		56	54		46	51	+5	45	43	
4 Sales	53	10.1	48	9.5	.38	−5	58	50	−8	50	44	−6	58	54	−4	58	49	−9	50	49		48	42	−6
5 Merchandising	50	9.9	50	10.4	.52		56	56		47	45		52	54		56	52	−4	47	50		46	43	
6 Office Practices	51	10.2	46	9.9	.56	−5	61	55	−6	50	43	−7	49	47		53	48	−5	47	45		46	42	−4
7 Military Activities	53	9.0	50	9.4	.43		53	48	−5	52	52		53	51		55	52		53	51		50	46	−4
8 Technical Supervision	48	9.6	50	11.7	.37		46	50	+4	48	47		46	42	−4	47	45		53	61	+8	46	48	
9 Mathematics	53	8.6	54	9.6	.69		54	57		53	53		44	41		49	50		59	60		57	59	
10 Science	52	9.5	53	9.3	.60		48	50		54	55		46	45		46	47		58	58		60	60	
11 Mechanical	48	9.9	48	10.9	.72		45	46		49	49		45	41	−4	42	42		57	58		45	47	
12 Nature	42	9.8	45	10.6	.64		38	39		44	49	+5	41	40		39	41		44	47		46	50	+4
13 Agriculture	43	10.0	44	10.1	.53		40	40		43	46		42	40		40	41		47	47		45	44	
14 Adventure	58	10.5	57	11.0	.47		56	53		59	56		57	59		59	56		62	61		55	53	
15 Recreational Leadership	52	9.8	50	9.8	.65		55	52		53	53		51	49		53	49	−4	51	53		46	47	
16 Medical Service	52	9.8	55	10.5	.56		48	50		57	63	+6	48	49		50	50		49	52		62	64	
17 Social Service	52	9.6	52	9.7	.44		55	54		50	50		53	56		53	53		48	49		55	51	−4
18 Religious Activities	52	10.4	53	11.3	.38		52	53		50	52		54	55		54	52		50	53		54	56	
19 Teaching	45	9.3	53	8.3	.36	+8	46	54	+8	42	52	+10	47	55	+8	46	53	+7	42	51	+9	46	55	+9
20 Music	46	10.3	51	9.4	.55	+6	48	51		44	48	+4	48	52	+4	50	53		41	49	+8	48	56	+8
21 Art	45	8.5	50	9.8	.57	+5	43	46		45	51	+6	48	55	+7	45	49	+4	44	49	+5	45	51	+6
22 Writing	49	9.2	53	10.1	.63	+4	50	41		43	45		58	64	+6	56	60	+4	44	49	+5	48	54	+6
Mdn					.56																			

[a] Mean changes larger than 4 points.

TABLE 58

THREE HIGHEST TEST AND RETEST MEANS AND LARGEST GAINS FOR BERDIE–SCHLETZER SIX CURRICULAR GROUPS

Accountants	*M*	Dentists	*M*	Journalists	*M*	Lawyers	*M*	Mechanical Engineers	*M*	Physicians	*M*
						Test					
Office Practices	61	Adventure	59	Writing	58	Public Speaking	63	Adventure	62	Medical Service	62
Sales	58	Medical Service	57	Sales	58	Law/Politics	61	Mathematics	59	Science	60
Merchandising	56	Science	54	Adventure	57	Adventure	59	Science	58	Mathematics	57
Adventure	56										
						Retest					
Law/Politics	60	Medical Service	65	Writing	64	Law/Politics	66	Technical Supervision	61	Medical Service	64
Mathematics	57	Adventure	56	Public Speaking	61	Public Speaking	65	Adventure	61	Science	60
Business Management	56	Science	55	Law/Politics	59	Writing	60	Mathematics	60	Public Speaking	59
Merchandising	56			Adventure	59					Mathematics	59
						Gains					
Teaching	46–54	Teaching	42–52	Teaching	47–55	Teaching	45–53	Teaching	42–51	Teaching	46–55
Law/Politics	55–60	Medical Service	57–63	Art	48–55	Law/Politics	61–66	Technical Supervision	53–61	Music	48–56
Technical Supervision	46–50	Art	45–51	Writing	58–64	Art	45–49	Music	41–49	Art	45–51
				Public Speaking	55–61	Writing	56–60			Writing	48–54

to measure specific areas of vocational interests. An important next step is to learn something about the application of these scales.

For Counseling

Counseling textbooks abound with viewpoints as to how test information should be used with individual clients. Research on these viewpoints is much less common, and there are very few data to support answers to even the most obvious questions: "It costs about one dollar to administer and score a student's Strong—is it worth it?" or "Should an individual be shown his scores?" In general, opinions here vary inversely in intensity with the amount of available relevant information.

For Selection

The situation with the use of interest inventories in employee selection is not greatly different—the basic questions on how to apply the methods have not been answered, and the only acceptable applications currently depend on the presence of a skilled, well-trained professional.

While much more research on the effects of various approaches is necessary, in the absence of such information the authors suggest the following guideline:

> These scales are reasonably accurate measures of an individual's interest in the specified areas. Anyone confronted with a choice of any of these activities, whether for training or employment, would be aided by this relevant information about himself. Professional psychologists faced with selection or placement decisions should find the results helpful in determining what activities an individual might find satisfying.

Though the basic scales should be useful in evaluating an individual's responses to the SVIB, they are intended to supplement, not supplant, the regular occupational scales. The regular scales still permit greater diversity and complexity in profile patterns than do the basic scales, as the unique items remaining in the SVIB, not included in any of the basic scales, are important. It would be a serious mistake to restrict measurements to these 22 factors in the belief that the valid variance in the interest domain has been covered, or even the valid variance in the SVIB item pool. Further research must push ahead to find other approaches, always with the intent of finding powerful yet parsimonious explanations of the organization of an individual's choices.

REFERENCES

BERDIE, R. F. Strong Vocational Interest Blank scores of high school seniors and their later occupational entry. *Journal of Applied Psychology*, 1960, 44, 161–165.

CAMPBELL, D. P. *Manual for the Strong Vocational Interest Blanks for men and women*. Stanford: Stanford University Press, 1966.

CLARK, K. E. *Vocational interests of non-professional men*. Minneapolis: University of Minnesota Press, 1961.

CRANNY, C. J. Factor analytically derived scales for the Strong Vocational Interest Blank. Unpublished doctoral dissertation, Iowa State University, 1967.

DARLEY, J. G., & HAGENAH, T. *Vocational interest measurement*. Minneapolis: University of Minnesota Press, 1955.

LAWLEY, D. N. The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburg*, 1940, **60**, 64–82.

SCHLETZER, V. M. A study of the predictive effectiveness of the Strong Vocational Interest Blank for job satisfaction. Unpublished doctoral dissertation, University of Minnesota, 1963.

STRONG, E. K., JR. *Change of interests with age*. Stanford: Stanford University Press, 1931.

STRONG, E. K., JR. *Vocational interests of men and women*. Stanford: Stanford University Press, 1943.

STRONG, E. K., JR. *Vocational interests 18 years after college*. Minneapolis: University of Minnesota Press, 1955.

THURSTONE, L. L. *Multiple factor analysis*. Chicago: University of Chicago Press, 1947.

WHERRY, R. J. Hierarchical factor solutions without rotation. *Psychometrika*, 1959, 24, 45–51.

VERBURG, W. A. Vocational interests of retired YMCA secretaries. *Journal of Applied Psychology*, 1952, 36, 254–256.

(Received January 2, 1968)

BUREAU OF EDUCATIONAL & PSYCHOLOGICAL RESEARCH CALCUTTA